R. Miller

JOURNAL OF

REGIONAL SCIENCE

Vol. 2 Fall 1960 No. 2

Richard Stone
A COMPARISON OF THE ECONOMIC STRUCTURE OF REGIONS BASED ON THE CONCEPT OF DISTANCE 1

John D. Herbert and Benjamin H. Stevens
A MODEL FOR THE DISTRIBUTION OF RESIDENTIAL ACTIVITY IN URBAN AREAS 21

Beverly Duncan and Otis Dudley Duncan
THE MEASUREMENT OF INTRA-CITY LOCATIONAL AND RESIDENTIAL PATTERNS 37

Josiah C. Russell
THE METROPOLITAN CITY REGION OF THE MIDDLE AGES 55

Stephen Spiegelglas
SOME ASPECTS OF STATE TO STATE COMMODITY FLOWS IN THE UNITED STATES 71

Michael F. Dacey
A NOTE ON THE DERIVATION OF NEAREST NEIGHBOR DISTANCES 81

PUBLISHED BY THE REGIONAL SCIENCE RESEARCH INSTITUTE
in cooperation with the
Department of Regional Science of the Wharton School
University of Pennsylvania

Vol. 2 Fall 1960 No. 2

JOURNAL OF REGIONAL SCIENCE

Edited by W. Isard and M.B. Teitz

Managing Editor, B.H. Stevens

The Journal of Regional Science is published in two or more issues per volume by the REGIONAL SCIENCE RESEARCH INSTITUTE, G.P.O. Box 8776, Philadelphia 1, Pa., in cooperation with the Department of Regional Science, Wharton School, University of Pennsylvania.

The subscription rate for libraries, business and other organizations, and individuals is $5.00 per volume. All orders should be addressed to the REGIONAL SCIENCE RESEARCH INSTITUTE and not to the Wharton School. *Individuals only* may receive a subscription by joining the Regional Science Association. Membership dues in the Association are currently $3.00 per year. For further information about membership in the Association, write to Regional Science Association, Wharton School, University of Pennsylvania, Philadelphia 4, Pa.

Back numbers and single issues may be obtained from the INSTITUTE while they are still in stock at a price of $2.50 per copy.

Manuscripts are invited for consideration for publication and should be sent, with other communications for the editors, to the REGIONAL SCIENCE RESEARCH INSTITUTE. All manuscripts should be in duplicate, typewritten, and double spaced. Mathematical symbols entered by hand must be clear. Figures should be drawn in black ink on white paper in a form suitable for photographic reproduction with the possibility of reduction taken into account. The style of the Journal should be followed with respect to footnotes and references, the latter being arranged in alphabetical order and grouped at the end of the article. Mention of particular works in footnotes should indicate the author's surname and number of the reference cited.

The REGIONAL SCIENCE RESEARCH INSTITUTE is a non-profit corporation devoted to research and studies on the structure, function, and operation of regions from an economic, social, and political standpoint.

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

A COMPARISON OF THE ECONOMIC STRUCTURE OF REGIONS BASED ON THE CONCEPT OF DISTANCE.

by Richard Stone*

1. INTRODUCTION

This paper is concerned with comparing the economic structure of regions on the assumption that we cannot define the concept of economic structure in concrete terms. Those readers who believe that this stage of regional analysis has already been passed can have no interest in what is proposed here and would be wasting their time in reading this paper: they begin where it leaves off, or beyond.

In the development and expansion of his great work *The Vectors of Mind*, Thurstone [16] writes:

> "The exploratory nature of factor analysis is often not understood. Factor analysis has its principal usefulness at the border line of science. It is naturally superseded by rational formulations in terms of the science involved. Factor analysis is useful, especially in those domains where basic and fruitful concepts are essentially lacking and where crucial experiments have been difficult to conceive. The new methods have a humble role. They enable us to make only the crudest first map of a new domain. But if we have scientific intuition and sufficient ingenuity, the rough factorial map of a new domain will enable us to proceed beyond the exploratory factorial stage to the more direct forms of psychological experimentation in the laboratory."

This paper is based on the concept of distance and the methods suggested for exploring regional structure are closely related to factor analysis. It is offered in the spirit of Thurstone's remarks where the 'rational formulations in terms of the science involved' either have not been made or are manifestly unsatisfactory.

---

*The author is Professor of Economics at the University of Cambridge, England.

The plan of this paper is as follows: section 2 is devoted to describing the problem of exploratory regional comparisons; section 3 introduces the data used; sections 4 and 5 outline the formal theory; section 6 sets out an application of the theory to the British economy; and section 7 contains some brief conclusions.

2. THE PROBLEM

If we divide an economy into regions it is easy to see that these regions have a number of characteristics which are not the same in every case. Some regions are rich, others poor; some are agricultural, others industrial; some are workplaces, others dormitories; some are exporters, others importers; and so on indefinitely. The first problem considered in this paper is: given a number of variables which may be used to characterize a region, is it possible to group the regions into clusters so that those in a given cluster are in some sense more like one another than they are like those in other clusters? The second problem is: if we can make these distinctions, can we say anything about the criteria implied in our analysis and can we translate these criteria into concrete terms?

FIGURE 1: The 12 Regions of the United Kingdom.
(1948 Population in '000's)

3. THE DATA

The method proposed in the following section could be applied to any set of economic measurements available for a number of different regions. In the applications given here, these measurements take the form of entries in a set of regional accounts. In [15] a set of 'national' accounts is given for each of the twelve civil defense regions of the United Kingdom in 1948. Each region is provided with a set of three accounts, relating respectively to production, consumption and accumulation. These are consolidated accounts and contain in all fourteen distinct transactions, only eleven of which are independent since each of the accounts must balance. The applications given below are based on these eleven variables for each of the twelve regions.

4. THE METHOD

This paper is based on the concept of Euclidean distance. We can set up eleven orthogonal axes corresponding to the eleven variables, and call the space so defined the transaction space. Each region can be represented by a point in transaction space. In all there are twelve such points each separated from the others and from the origin of the space by a certain distance.

To put the problem on a more general footing, consider n regions, $j, k = 1, 2, \ldots, n$, and m independent transactions, $r, s = 1, 2, \ldots, m$. Then from standard works on n-dimensional geometry we know that the square of the distance of region j from the origin of the space, d^2_{jo} say, is given by

$$(1) \qquad d^2_{jo} \equiv \sum_r x^2_{rj} \equiv x'_j \, x_j$$

where x_{rj} is the value of transaction r in region j, the vector $x_j \equiv \{x_{1j}, \ldots, x_{mj}\}$ and x'_j denotes the transpose of x_j. The square of the distance between two regions, j and k, is given by

$$(2) \qquad d^2_{jk} \equiv \sum_r (x_{rj} - x_{rk})^2$$

$$\equiv (x_j - x_k)' \, (x_j - x_k).$$

As they stand, these definitions of distance are unsatisfactory. First, if one region is much larger, in population for example, than the others, it is likely that most, if not all, of its transactions will also be much larger than their counterparts in other regions. Accordingly this large region will appear as very distant from the others simply because it is large. To avoid this result we must adjust the regions for size. One way to do this would be to divide the transactions of each region by that region's population. If we denote the population of region j by π_j, then we should replace the vector x_j by x^*_j where

$$(3) \qquad x^*_j \equiv \pi_j^{-1} \, x_j$$

Evidently there is nothing compelling about the use of population in this way. In (3) regions will no longer be distinguished by size, as measured by population, but rich regions, with a high level of production per head, will still tend to be distant from poor ones. If we wished to remove this distinction we should have to divide the regional transactions not by the region's population but by, say, its gross domestic product.

In fact each transaction of a region need not necessarily be divided by the same magnitude. Thus we might divide the individual outgoings of each account by that account's total outgoings (≡ total incomings). If this method were adopted the variables to be analysed would be similar to those in a matrix of input-output coefficients.

The point that emerges from this discussion is that as we change the variable used to standardize the observations we change the characteristics of the regions which contribute to the distance between them.

But this is not the only respect in which the original definition of distance is unsatisfactory. For, second, it may not seem sensible to allow distance to depend on the relative size of transactions. For example domestic product per head is a large transaction capable of large absolute variation from region to region. By comparison, saving per head is a small transaction showing, typically, a much smaller variation. In defining distance we may not wish to give one transaction more weight than another. If this is so we should be led to normalizing the transactions by making their sum of squares over the regions equal to *1* in each case. Thus if x^*_{rj} is the standardized value of transaction r in region j, we might convert it into a normalized measure by putting

$$\sum_j (x^*_{rj})^2 = 1 \tag{4}$$

so that the normalized measures, x^{**}_{rj} say, are given by

$$x^{**}_{rj} \equiv \frac{x^*_{rj}}{\sum_j (x^*_{rj})^2} \tag{5}$$

or, in vector form,

$$x^{**}_j \equiv \hat{q}^{-1} x^*_j \tag{6}$$

where the rth element of the diagonal matrix $\hat{q}$ is $\sum_j x^{*2}_{rj}$.

Finally, there is a third consideration to be taken into account. Even if we standardize and normalize the observations we have still made no allowance for the fact that the different variables used to characterize the regions are likely to be quite highly correlated. If each pair of transactions was perfectly correlated over the regions we should gain nothing by looking at more than one, since each would tell the same story. We can allow for the intercorrelations of the variables by reducing them to a set of hypothetical orthogonal variables in terms of which distance can be measured by means of (1) and (2) since they can properly be referred to a set of orthogonal axes.

If the variables (the transactions) are linearly independent then they are connected to the hypothetical orthogonal transactions, z_{rj} say, by the linear transformation

$$x_j^{**} = A\, z_j \tag{7}$$

Where A is a non-singular matrix of order m. If we now calculate distance in terms of the z's we see that

$$d_{jo}^2 \equiv z_j'\, z_j \equiv x_j^{**\prime}(AA')^{-1}\, x_j^{**} \tag{8}$$

$$\equiv x_j^{**\prime}\, R^{-1}\, x_j^{**}$$

where R is the matrix of zero-order correlations between the transactions. In a similar way (2) becomes

$$d_{jk}^2 \equiv (x_j^{**} - x_k^{**})'\, R^{-1}\, (x_j^{**} - x_k^{**}) \tag{9}$$

which expresses the square of the distance between two regions, j and k, as a quadratic form in the differences of the observed values of their transactions, each expressed in standard, normal form.

From the observations we can calculate the elements of R and of the x_j^{**}. Accordingly from (9) we can calculate the $\frac{n(n-1)}{2}$ distances between the n regions. We may then attempt to group the regions into clusters in such a way that the average value of d^2 within clusters is small compared with the average value of d^2 between clusters. This problem is discussed by Rao in [11]. He points out that grouping is an arbitrary process, and that different definitions of the qualities distinguishing the members of one group from those of another will lead to differing procedures. This point will be taken up more fully in section 6 of this paper.

5. DISTANCES AND FACTORS

Equation (9) will be recognized as formally similar to Mahalanobis' generalized distance set out many years ago in [7, 8, 9, 10]. In these applications, races, castes or tribes corresponded to regions and anthropometric measurements to transactions. An important difference which distinguishes these from the present application is that Mahalanobis had

available a number of sets of measurements for each race, whereas in the regional analysis attempted here there is only a single set of measurements for each region. Consequently Mahalanobis was able to obtain measures of the correlations between characteristics by pooling the correlations obtained within each racial group whereas in the present case the correlations are obtained by correlating pairs of transactions over regions. The statistical aspects of generalized distance have been discussed in detail by R. C. Bose in [1, 2], S. N. Bose in [3, 4], R. C. Bose and S. N. Roy in [5], and Roy in [12, 13].

Again, the matrix A in (7) will be recognized as the pattern matrix of factor analysis and the relationship $R = AA'$ of which use is made in (8) as what is sometimes called the fundamental theorem of factor analysis. In (7) the elements of z_j can be interpreted as a set of m orthogonal hypothetical transactions or factors in terms of which the actual transactions can be described. The matrix A contains in row s the weights necessary to construct transaction s as a linear combination of the hypothetical factors. In Hotelling's method of principal components [6], the first component or factor is chosen in such a way as to remove as large a part as possible of the sum of squares of the observations, that is to minimize the expression

$$\sum_r \sum_j (x_{rj} - a_{r1} f_{1j})^2$$

where f_{1j} is the value of the first factor in region j and a_{r1} is the element in row r and column 1 of A. The second factor is then chosen so as to remove as large a part as possible of the residual sum of squares left when the variation attributable to the first factor has been removed. Provided that the transactions are linearly independent this process, if continued, will eventually lead to m orthogonal factors or hypothetical transactions in terms of which the whole of the original sum of squares of the observations can be accounted for. These factors are the characteristic vectors associated with the correlation matrix, R, the roots of whose characteristic equation are all necessarily real and positive.

One of the interesting features of factor analysis is that, with correlated variables, it is usually possible to account for most of the original sum of squares with far fewer factors than variables. This situation can be described geometrically. Suppose there are three regions and any number of transactions. If the regions are represented by three rectangular axes and if the transactions are normalized, then each transaction can be represented as a point in the region space which is a unit distance from the centre of the coordinate system. Thus each transaction can be represented as a point on the unit sphere. If the third orthogonal factor accounts for a negligible amount of the total sum of squares, then the points representing the transactions will all lie close to a great circle drawn on the surface of the sphere. If both second and third orthogonal factors are unimportant then all the transactions will lie in the neighborhood of a single point on the surface of the sphere.

Distance analysis and factor analysis are thus closely related: a distance analysis for regions is, in a sense, complementary to a factor analysis for transactions. But the roles of regions and transactions could be interchanged: we could imagine a factor analysis for regions and a complementary distance analysis for transactions.

An economic application of Hotelling's method is given in [14]; the variables were again transactions but the observations related to years rather than regions. The method described is, however, exactly the same.

6. APPLICATIONS

As mentioned above in section 2, the theory of the preceding section will now be applied to data for the economic activity of the United Kingdom during the year 1948. The twelve civil defense regions of the United Kingdom are shown with their population in 1948 in the sketch-map. The accounting data are based entirely on the unpublished work of Phyllis Deane and are described in more detail in section 6 of the present writer's earlier paper on regional accounting [15]. In Table 1 the actual figures for the eleven transactions used are set out. These are taken from a complete set of fourteen transactions, from which three are omitted corresponding to three accounting identities which hold between the fourteen transactions. The symbols given in Table 1 are the same as those used in [15], and the data are given in millions of pounds sterling.

The analyses performed on the data of Table 1 fall into three classes. In each analysis, the figures for each region have been divided by the regional population in order to remove a gross size effect, as in (3) above. As before, the variable x^*_{rj} denotes the value of the r'th transaction in the j'th region, and the vector x^*_j denotes the vector of all transactions for the j'th region.

In analysis A the distances between each pair of regions, j and k say, are measured by (2), repeated here as

$$(10) \qquad d^2_{jk}\,(\mathrm{A}) \equiv (x^*_j - x^*_k)' \;\; (x^*_j - x^*_k)$$

Thus in this analysis the regions are represented by points in 11-dimensional space defined by eleven orthogonal axes corresponding to the transactions.

In analysis B the transactions have been first separately normalized *within* each region so that

$$(11) \qquad d^2_{jk}\,(\mathrm{B}) \equiv (x^{**}_j - x^{**}_k)' \;\; (x^{**}_j - x^{**}_k)$$

In this analysis each region is represented as a point on the unit hypersphere in 11-dimensional space defined again by eleven orthogonal axes corresponding to the transactions.

TABLE 1: Accounting Data Used in the Analysis

Transaction No.	1	2	3	4	5	6	7	8	9	10	11
Transaction Title	Consumption	Gross investment	Gross domestic product	Gross saving	Indirect taxes	Direct taxes	Central government borrowing	Sales to central government	Current transfers from central government	Capital transfers from central government	Current transfers from the rest of the world
Transaction Symbol	C_{jj}	V_{jj}	Y_{dj}	S_{jj}	I_{gj}	D_{gj}	B_{gj}	C_{jg}	G_{jg}	T_{jg}	G_{jr}
1. North	482	106	520	51	68	95	-63	59	77	-1	0
2. E. and W. Ridings	690	171	844	105	103	167	-79	72	97	2	29
3. N.Midlands	551	127	620	72	85	126	-71	66	74	-2	7
4. East	498	91	489	65	88	90	-53	69	74	-3	-29
5. London and S.E.	2348	393	3096	440	499	566	-269	311	306	20	21
6. South	430	75	379	37	76	74	-56	108	68	-3	-32
7. South West	451	89	465	56	71	88	-61	97	73	-3	-49
8. Wales	364	81	377	36	32	67	-45	44	71	0	-34
9. Midlands	719	179	908	103	139	178	-90	91	96	2	-23
10. North West	1111	282	1300	153	186	253	-128	126	161	7	46
11. Scotland	878	173	947	113	97	169	-97	81	127	-11	44
12. N.Ireland	171	30	178	19	5	31	-16	14	26	0	-15

NOTE: The transactions omitted because of the adding-up restraints are

(i) Net exports to the rest of the world (X_{jr}) $\equiv 3 + 5 - 1 - 2 - 8$

(ii) Net factor income from the rest of the world (Y_{jr}) $\equiv 1 + 4 + 6 - 3 - 9 - 11$

(iii) Borrowing by the rest of the world (B_{rj}) $\equiv 4 + 10 - 2 - 7$

In analysis C the transactions have been normalized *over* regions, and account has been taken of the correlations between transactions so that

$$(12) \qquad d^2_{jk}\ (C) \equiv (x^{**}_j - x^{**}_k)'\ R^{-1}\ (x^{**}_j - x^{**}_k)$$

This is similar to the analysis A except that it is not assumed that the transactions lie on orthogonal axes. For a reason which will be discussed later, this analysis (C) has been carried out on the first three, the first six and the first nine transactions of Table 1.

In each analysis the $\frac{12 \times 11}{2} = 66$ elements of the distance matrix have been calculated, and attempts made to group the regions into some plausible classification. The purpose of grouping is clearly to place within the same group regions whose mutual distances are small, but as Rao [11] has pointed out, the exact method for grouping will depend on making the concept of a group more precise, as can be done in a number of arbitrary ways. Thus the method will differ, for example, according to whether simple clusters, rings or chains are looked for. The technique adopted is suitable for discovering simple clusters. First, two measures are defined, a measure of the compactness of a group of regions, and a measure of distance between groups of regions; these definitions are based on the idea of average distances. Let j and k be typical regions in group g, containing p regions, and let u and v be typical regions in group h, containing q regions. Then the measures of compactness are given by

$$(13) \qquad D^2_{gg} \equiv \frac{2}{p(p-1)} \sum_{j<k} d^2_{jk}$$

$$(14) \qquad D^2_{hh} \equiv \frac{2}{q(q-1)} \sum_{u<v} d^2_{uv}$$

and the distance between group g and group h by

$$(15) \qquad D^2_{gh} \equiv \frac{1}{pq} \sum_{j=1}^{p} \sum_{u=1}^{q} d^2_{ju}$$

Beginning with all regions unclassified, the two with minimum distance are chosen; these form the nucleus of the first group with $D^2_{11} \equiv d^2_{jk}$, if j and k are these two regions. The remaining regions are then ordered according to their average distance from the regions in group *1*. The region with the minimum average distance is then included in this group if its average distance is not greater than say αD^2_{11}, where α is an arbitrary constant. In this case D^2_{11} is recalculated for the three regions and the procedure repeated. Otherwise the third region forms the nucleus of a new group. The next nearest region to the third is included with it if its distance from it is less than αD^2_{22}, and the procedure repeated.

This procedure is one in which the criterion of inclusion changes through the analysis, as more regions are included in groups or new groups formed, and may be called "grouping according to the sequential criterion". Another method has also been tried in which the mean of all the elements of the distance table is first calculated, $\bar{d}^2$, say, and regions are classified together if d^2_{jk} say, is not greater than $\alpha\bar{d}^2$. This procedure may be called "grouping according to the absolute criterion."

The results for analysis A are given in Tables 2 and 3. In Table 2 the values of d_{jk}(A) are given, and in Table 3 the grouping, intra-group and inter-group distances are shown. The differences between the two methods are not very great, being mainly a matter of the demarcation line between groups II and III. In fact, it is possible to represent analysis A approximately on a line diagram, in which the regions are ordered as they appear in Table 3, with London heading the list and Northern Ireland at the bottom. This is so because from Table 3 it can be seen that

$$D_{gh} \propto D_{gi} + D_{ih} \tag{16}$$

for any g, h, i, provided the elements of the matrix are appropriately signed. The ordering arrived at is closely related to gross domestic product per person, and is therefore roughly equivalent to an ordering of the regions by the criterion of productiveness.

The results of analysis B are set out in Tables 4 and 5. The normalization of the transactions within regions, that is the division of each transaction for a region by the sum of squares of all transactions for the same region, removes any factor which is consistently greater in one region than another over all transactions, and therefore places more emphasis on the structure of transactions, that is the size of the individual transactions in relation to each other. Once again the effect of using different grouping criteria is somewhat trivial, merely changing the demarcation line, as it were, between groups in a fairly simple way.

With regard to analysis C there are certain limitations which are characteristic of the present application. The method of analysis C is suggested, as described in section 5 above, by analogy with factor analysis, which was developed within the field of psychometrics. In intelligence tests there are usually no common units of measure between the variables, which are typically the scores of students when examined say in mathematics or in languages. Further, there are usually sufficient examination results to enable the correlation matrix between variables to be estimated reasonably well. In the present application to regional accounting we have, on the other hand, a common unit of account, in which the different transactions are measured while the number of regions is very little greater than the number of transactions. If the number of regions is equal to (or less than) the number of transactions, the correlation matrix as estimated from the data has no inverse and d^2_{jk}(C) cannot be calculated. The reason for this is that if the n points representing the regions are plotted in n dimensional

TABLE 2: Interregional Distances, Analysis A

Regions	Regions											
	1	2	3	4	5	6	7	8	9	10	11	12
1. North	-	14.5	8.1	6.1	44.4	11.8	7.8	9.9	14.8	13.9	8.3	17.3
2. E. and W. Ridings		-	6.9	15.8	30.9	21.8	19.3	23.7	4.9	2.7	8.5	30.7
3. N. Midlands			-	9.3	36.7	15.5	12.9	17.4	7.7	6.3	4.5	24.8
4. East				-	43.5	8.3	6.3	11.9	15.2	14.7	9.9	19.5
5. London and S.E.					-	47.7	47.0	53.2	30.8	31.3	37.7	60.5
6. South						-	7.2	13.8	21.0	20.4	16.2	20.9
7. South West							-	8.9	18.1	18.0	14.1	16.7
8. Wales								-	23.3	23.3	17.4	8.8
9. Midlands									-	5.4	10.8	30.6
10. North West										-	7.9	30.7
11. Scotland											-	24.3
12. N. Ireland												-

TABLE 3. Groupings of Regions in Analysis A by Alternative Criteria

Criterion 1. Absolute Criterion

Regional Groupings

I London and S.E.

II E. and W. Ridings, N. Midlands, Midlands, North West, Scotland.

III North, East, South, South West, Wales.

IV N. Ireland.

Distances within and between groups

	I	II	III	IV
I	0	34	46	61
II		7	17	28
III			10	17
IV				0

Criterion 2. Sequential Criterion

Regional Groupings

I London and S.E.

II E. and W. Ridings, North West, Midlands, Scotland.

III North, N. Midlands, East, South, South West, Wales

IV N. Ireland

Distances within and between groups

	I	II	III	IV
I	0	31	43	61
II		2	17	31
III			13	21
IV				0

Note: In the distance table the diagonal elements indicate the average distances between regions within a group, and the off-diagonal elements the distances between groups.

TABLE 4: Interregional Distances, Analysis B

Regions	Regions											
	1	2	3	4	5	6	7	8	9	10	11	12
1. North	-	24.6	11.8	23.7	38.9	48.7	31.9	24.8	28.0	18.6	16.2	33.2
2. E. and W.Ridings		-	15.4	43.2	24.9	69.6	49.3	43.4	17.5	9.2	22.7	45.5
3. N. Midlands			-	29.6	31.3	55.6	37.0	33.0	20.3	10.2	17.0	38.5
4. East				-	49.2	33.5	21.5	26.9	41.0	37.2	33.2	36.7
5. London and S.E.					-	74.4	52.6	54.7	20.9	29.1	38.9	55.6
6. South						-	28.6	46.0	66.4	62.9	59.2	58.8
7. South West							-	27.0	42.8	44.6	44.6	40.5
8. Wales								-	41.8	39.9	34.9	21.1
9. Midlands									-	19.2	33.9	47.0
10. North West										-	19.9	45.1
11. Scotland											-	34.8
12. N. Ireland												-

TABLE 5: Grouping of Regions in Analysis B by Alternative Criteria

Criterion 1. Absolute Criterion

Regional Groupings

I: London and South East

II: Midlands, E. and W. Ridings, North West, North, N. Midlands, Scotland

III: East, South West

IV: Wales, N. Ireland

V: South

Distances within and between groups

	I	II	III	IV	V
I	0	3.14	5.09	5.51	7.44
II		2.00	3.88	3.89	6.08
III			2.15	3.33	3.11
IV				2.11	5.28
V					0

Criterion 2. Sequential Criterion

Regional Groupings

I: London and South East, Midlands

II: E. and W. Ridings, North West

III: North, N. Midlands, Scotland

IV: East, South West, Wales, N. Ireland

V: South

Distances within and between groups

	I	II	III	IV	V
I	2.09	2.31	3.25	4.84	7.05
II		0.92	1.92	4.36	6.63
III			1.52	3.37	5.46
IV				2.99	4.34
V					0

space they define a figure which could be plotted in $(n-1)$ dimensional space, at most, from which it is not possible to estimate the parameters which could transform the n-space into a different n-space. If there is one more region than the number of transactions the calculation is possible but extremely restricted. It is obvious that with 2 regions and 1 transaction the normalized distance between the regions always reduces to the same numerical value. And in the present application with 12 regions and 11 accounts the interregional distances will reduce to

(17) $$d^2_{jk}\ (\mathrm{C}) = 2400.$$

Accordingly the analysis has been applied with 3, 6 and 9 transactions, each taken in order from Table 1. The distance matrix and groupings for the 3-transaction analysis are given in Tables 6 and 7.

TABLE 6: Interregional Distances, Analysis C
Three Transactions

Regions	Regions											
	1	2	3	4	5	6	7	8	9	10	11	12
1. North	-	13.7	7.7	15.2	36.9	25.5	7.1	7.2	15.3	17.2	10.1	26.4
2. N. and W. Ridings		-	8.9	25.6	36.3	34.6	20.2	18.4	6.0	9.4	18.3	35.6
3. N.Midlands			-	17.2	36.2	25.9	13.5	14.3	13.2	10.0	10.8	33.6
4. East				-	35.5	11.7	11.4	19.4	28.6	25.3	8.1	32.7
5. London and S.E.					-	43.4	35.5	41.4	35.9	41.0	30.9	43.8
6. South						-	22.5	29.3	38.6	31.5	18.8	42.8
7. South West							-	8.9	21.2	23.4	8.9	23.0
8. Wales								-	18.3	22.8	16.1	21.2
9. Midlands									-	15.2	21.1	32.7
10. North West										-	19.7	42.5
11. Scotland											-	30.9
12. N.Ireland												-

TABLE 7: Groupings of Regions in Analysis C
Three Transactions, Absolute Criterion.

Regional groupings

I London and S.E.

II E. and W. Ridings, North Midlands, Midlands, N. West.

III North, Scotland, South West, Wales

IV East, South

V N. Ireland.

Distances within and between groups

	I	II	III	IV	V
I	0	37.4	36.4	39.6	43.8
II		10.9	17.8	29.1	36.3
III			10.1	19.9	25.6
IV				11.7	38.1
V					0

Tables 8 and 9 show the distance matrix for 6 and 9 transactions respectively. It will be noticed that in the 9-transactions table the elements are approaching $49 = \sqrt{2400}$ in each case.

TABLE 8: Interregional Distances, Analysis C

Six Transactions

Regions	Regions											
	1	2	3	4	5	6	7	8	9	10	11	12
1. North	-	36.5	37.9	39.4	40.3	34.5	30.2	13.6	24.6	26.9	32.3	38.1
2. E. and W. Ridings		-	17.2	37.6	40.5	43.5	28.6	32.1	32.1	19.8	31.1	36.6
3. N. Midlands			-	42.2	42.8	32.6	29.6	36.3	33.8	29.8	36.5	35.8
4. East				-	44.5	43.1	19.1	36.3	42.8	33.2	39.6	44.4
5. London and S.E.					-	46.6	39.5	43.1	39.9	42.2	39.7	45.4
6. South						-	35.8	39.0	39.4	41.6	40.6	46.6
7. South West							-	27.8	28.8	29.8	38.3	30.3
8. Wales								-	30.7	25.5	25.8	30.3
9. Midlands									-	29.6	47.0	41.6
10. North West										-	27.9	45.4
11. Scotland											-	39.5
12. N. Ireland												-

TABLE 9: Interregional Distances, Analysis C

Nine Transactions

Regions	Regions											
	1	2	3	4	5	6	7	8	9	10	11	12
1. North	-	44.9	39.7	43.6	44.6	44.3	43.9	36.9	31.0	38.6	35.9	45.1
2. E. and W. Ridings		-	35.1	44.3	41.4	44.2	41.5	37.9	37.7	23.0	40.4	40.4
3. N. Midlands			-	47.0	48.3	47.2	48.3	47.7	44.4	46.3	45.6	48.1
4. East				-	48.3	49.0	48.5	47.7	46.4	40.7	47.7	47.6
5. London and S.E.					-	48.5	49.0	47.1	43.9	45.3	45.7	48.9
6. South						-	48.7	47.6	45.8	41.8	47.2	48.1
7. South West							-	47.6	44.8	44.7	46.4	48.7
8. Wales								-	48.1	40.7	48.5	46.3
9. Midlands									-	32.2	48.8	42.2
10. North West										-	35.5	46.3
11. Scotland											-	44.4
12. N. Ireland												-

7. CONCLUSIONS

In this paper several simple methods are presented for comparing the economic structure of a number of regions as revealed by their transactions in a given year.

From a technical point of view analysis A suffers because, though a gross size factor has been removed at the outset by expressing all variables per head of population, the results are likely to be dominated by any remaining tendency which causes all the transactions of one region to be systematically greater than those of another.

Analysis B, whilst being only slightly more difficult to compute than analysis A, removes this further factor by deflating the figures for each region by the sum of the squares of all transactions for the region. Since this is done separately for each region, problems of comparing of the unit of account in different regions, as would arise in comparing different countries within the same geographical area, are avoided. Though the analysis does not formally take into account the possibility that the different transactions may be inter-correlated, this is largely revealed by the analysis, for it may be easily tested whether or not the results can be approximately shown on a line or plane diagram. Such analysis may help in the decision as to whether a formal factor analysis would be worth while.

Analysis C, which has a sounder theoretical base than the preceding analyses, leads to difficulties if the number of regions does not greatly exceed the number of transactions unless it is possible to replace the actual transactions by a relatively small number of hypothetical transactions or factors. To carry out the analysis it must be assumed that there is a common unit of account between regions, though not necessarily between transactions. Indeed any original unit of account between transactions is destroyed in the process of the analysis, which reduces all transactions to scales on which they show equal variability between regions. The number of calculations required in the analysis is very much greater than in analyses A and B.

In the present application the qualitative results of all the analyses are very similar. London and South-East region is usually clearly separated from the rest. Next come the regions which contain the other great conurbations and industrial centers of the United Kingdom, the East and West Ridings, North West, Midlands, North Midlands and Scotland. The East, South West and Wales usually form a third group. Northern Ireland is usually clearly separated, and so also is the South region, which has no specific character, ranging from the South Coast, with Southampton, Portsmouth and holiday resorts, through sparsely populated agricultural counties, to as far north as Oxford.

REFERENCES

[1] Bose, R.C. 'On the Exact Distribution and Moment-Coefficients of D^2-Statistics' in *Sankhyā*, Vol. 2, pt. 2, 1936, pp. 143-54.

[2] Bose, R.C. 'A Note on the Distribution of Differences in Mean Values of Two Samples Drawn From Two Multivariate Normally Distributed Populations, and the Definition of the D^2 - Statistic' in *Sankhyā*, Vol. 2, pt. 4, 1936, pp. 379-84.

[3] Bose, S.N. 'On the Complete Moment-Coefficients of the D^2 -Statistic' in *Sankhyā*, Vol. 2, pt. 4, 1936, pp. 385-96.

[4] Bose, S.N. 'On the Moment-Coefficients of the D^2 - Statistic and Certain Integral and Differential Equations Connected with the Multivariate Normal Population' in *Sankhyā*, Vol. 3, pt. 3, 1937, pp. 105-24.

[5] Bose, R.C., and S.N. Roy, 'The Distribution of Studentised D^2 - Statistic' in *Sankhyā*, Vol. 4, pt. 1, 1938, pp. 19-38.

[6] Hotelling, Harold. 'Analysis of a Complex of Statistical Variables into Principal Components,' in *Journal of Educational Psychology*, Vol. XXIV, 1933, pp. 417-41 and 498-520.

[7] Mahalanobis, P.C. 'Analysis of Race Mixture in Bengal' in *Journal and Proceedings of the Asiatic Society of Bengal*, Vol. XXIII (new series), No. 3, 1927, pp. 301-33.

[8] Mahalanobis, P.C. 'On Tests and Measures of Group Divergence' in *Journal of the Asiatic Society of Bengal*, Vol. XXVI (new series), No. 4, 1930, pp. 541-88.

[9] Mahalanobis, P.C. 'Anthropological Observations on the Anglo-Indians of Calcutta. Statistical Analysis of Measurements of Seven Characters' in *Records of the Indian Museum*, Vol. XXIII, Pt. 111, 1940, pp. 151-87.

[10] Mahalanobis, P.C. 'Discussion on the Application of Statistical Methods in Anthropometry' in *Sankhyā*, Vol. 4, pt. 4, 1940, pp. 594-8.

[11] Rao, C.R. *Advanced Statistical Methods in Biometric Research*, (Wiley, New York, 1952).

[12] Roy, S.N. 'A Note on the Distribution of the Studentised D^2 - Statistic' in *Sankhyā*, Vol. 4, pt. 3, 1939, pp. 373-80.

[13] Roy, S.N. 'p-Statistics or Some Generalisations in Analysis of Variance Appropriate to Multivariate Problems' in *Sankhyā*, Vol. 4, pt. 3, 1939, pp. 381-96.

[14] Stone, Richard. 'On the Interdependence of Blocks of Transactions' in *Supplement to the Journal of the Royal Statistical Society* Vol. IX, nos. 1-2, 1947, pp. 1-32.

[15] Stone, Richard. 'Social Accounts at the Regional Level: A Survey' (E.P.A. Conference, Bellagio, June, 1960, to be published).

[16] Thurstone, L.L., *Multiple-Factor Analysis* (Chicago University Press, 1947).

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

A MODEL FOR THE DISTRIBUTION OF RESIDENTIAL ACTIVITY IN URBAN AREAS.†

by John D. Herbert and Benjamin H. Stevens*

1. INTRODUCTION

The model presented here is designed to distribute households to residential land in an optimal configuration. The model was constructed for the Penn-Jersey Transportation Study as part of a larger model designed to locate all types of land-using activity.

Since the model had to be suitable for practical application a certain amount of conceptual elegance has been sacrificed in favor of operational simplicity. The larger model operates in the following way: The total relevant time period is subdivided into a number of short iterative periods. For each iterative period different types of land-using activity are handled separately. A particular type of activity is distributed in a configuration that is optimal only with respect to all previously located activities.[1] Interactions that are expected to occur among simultaneously-locating activities are ignored.

---

†This paper is adapted from a report prepared for the Penn-Jersey Transportation Study. The authors wish to acknowledge the contributions of Britton Harris, who directed the section of the Study under which the model was prepared and who made invaluable criticisms and suggestions throughout its development, Nan Fetter, and the other members of Penn-Jersey staff who participated in discussions of the problem. The paper also draws on the excellent theoretical development of similar materials in Alonso [1]. The authors accept full responsibility for the model's shortcomings.

*John D. Herbert is a doctoral student in the Department of City Planning at the University of Pennsylvania. Benjamin H. Stevens is Assistant Professor of Regional Science at the University of Pennsylvania.

[1]For a particular type of activity in a particular iterative period, previously located activities comprise all activities located in previous periods plus activities of other types previously located (by other elements in the larger model) in the same period.

We are assuming that they can be ignored if iterative periods are kept short enough to ensure that the number of users located in a single run of the model is small. Operating in this way we are able to achieve computational simplicity and, at the same time, recognize most of the basic interactions among land users.[2]

For the residential model, in a particular iterative period, the number of households to be located and the amount of land that is expected to be available for residential use is forecast exogenously.[3] A linear program is used to produce, for the end of that period, an optimal configuration of the new households on the available land. This configuration is optimal with respect to the configuration of all previously located activities, and constitutes a prediction of the way in which the forecast households will locate.[4]

2. DEFINITIONS

Household: A household is a person or group of persons with a common budget purchasing a single residential bundle.

Household Group: A household group is a collection of households which have similar residential budgets and similar tastes with respect to residential bundles.[5]

---

[2]For example, if we take an extreme case with an iterative period of one week, the number of land users that will be located in that period is likely to be small; it seems reasonable, both conceptually and realistically, to assume that they will make their locational decisions largely independent of one another. However, interactions between users located in a particular week and those located in previous weeks will be recognized, with the result that a vast majority of the important interactions are taken into account. For Penn-Jersey we envisage an iterative period of at least a year, which will certainly introduce inaccuracies; but we can achieve any level of accuracy that we desire, at the cost of increasing computational complexity, by decreasing the length of the periods.

[3]The residential model can handle land that is vacant, partially improved, or completely built-up. For expositional simplicity, the discussion will be limited to vacant land unless otherwise noted. Forecasts are exogenous in the sense that they are made outside the overall model (by techniques that are beyond the scope of the present discussion) but can be modified for a particular iterative period to recognize the configurations produced by the model in previous periods.

[4]Linear programming is not ordinarily regarded as a predictive tool. However, if we have a prediction of the number of households that is to be located and the model locates them in a realistic configuration then we can use the model to predict configurations. Since the configurations it produces are optimal in a specific economic sense the model may be both predictive and prescriptive.

[5]We are assuming that households that have similar tastes give rise to similar phenomena on the sites they occupy. E.g., households with similar tastes with respect to travel patterns generate the same numbers of each kind of trip, regardless of the number of persons in the household.

House: A house is the physical structure to be occupied by a single household.

House Cost: House cost is the dollar cost of constructing, operating and maintaining a house over a specified time period, computed on an annual basis.[6]

Amenity Level: The amenity level associated with a site is the level of psychic satisfaction which a household has an opportunity to enjoy because of certain characteristics of that site.[7]

Amenity Cost: The amenity cost associated with a particular site is the annual dollar cost of providing a specified amenity level on that site.[8]

Trip Set: A trip set consists of the numbers of each type of trip generated per year by a household.[9]

---

[6]We hope to obtain house costs from contractors' estimates. Clearly, the construction, operating and maintenance costs for a particular type of house may vary from area to area, and will be dependent in part upon the topography, soil, and micro-climate in each area.

[7]The site characteristics contributing to the amenity of a site have yet to be selected. It is expected that they will include characteristics such as the existing households in the area in which the site is located, the levels of public services in the area, the age of the area, other land uses in the area, planting on the site, and views from the site.

[8]Tentatively, it is proposed that estimates of amenity costs be obtained in the following way: Realtors and developers will be asked to identify for each area in the region the "premiums" they could add to or subtract from the selling price of a particular type of house, offered to a particular type of household because of the amenity of that area. Where these premiums reflect the accessibility of an area as well as its amenity level, we anticipate that the effects of accessibility can be eliminated. For a particular type of household and house the area in which that premium is a maximum will be regarded as having zero amenity cost for that type of household purchasing the residential bundle in which that type of house occurs. The corresponding amenity cost in any other area will be the difference between the premium associated with it and the maximum premium. Since premiums may be negative, amenity costs in some areas may be so high that they prevent the model from locating particular types of households in those areas. It is worth noting here that the model can recognize, in a similar way, segregation and other policies which affect the prices households are willing to offer for houses.

[9]By kind of trip we mean, for example, a trip to a particular type of employment, recreation, shopping, etc. The origins and destinations of the trips in a set are not specified since we assume that the alternative trip sets considered by a household will be independent of that household's location. There is some evidence in support of this assumption. See Marble [7]. It will be seen below that the assumption may not be unreasonably restrictive since we provide households with a number of alternative trip sets from which to choose.

Trip Pattern: The trips in a trip set when their origin and destinations have been identified.[10]

Travel Cost: Travel cost is the annual dollar cost to the household of carrying out a trip pattern.[11]

Site: A site is the parcel of land assignable to a particular household.[12]

Total Site Rent: Total site rent is the dollar value of the amount received annually by a site owner for the use of the *land* in the site. It is exclusive of the value of the house, amenity level and travel pattern associated with the site.

Unit Site Rent: The unit site rent on a site is the total site rent divided by the number of acres in the site.

Residential Bundle: A residential bundle is a unique combination of a house, an amenity level, a trip set, and a site of a particular size.

Market Basket: A market basket is a unique combination of a residential bundle and a bundle of all other commodities (which we shall call the "other commodities" bundle) consumed annually by a household.

Total Household Budget: A household's total budget is the dollar amount that a household allocates annually to the purchase of a market basket.

Residential Budget: A residential budget is that part of a household's total budget that is allocated annually to the purchase of a residential bundle.

Region: A region is the geographical space within which our model is required to allocate a given number of households to a given supply of land.

Area: An area is a regional subdivision whose characteristics are homogeneous with respect to the costs of construction, amenity costs, and the costs of transportation to other areas.

---

[10] The destinations of the trips in a trip set will vary with the location of the household generating the set. These destinations will be forecast exogenously by a probability interaction model. Cf. Carroll and Bevis [2].

[11] Travel costs, including out-of-pocket and imputed time costs, can be estimated from data on the transportation system.

[12] Land not covered by a house but under the same tenancy or ownership as the house is included in the site associated with that house. Where a parcel of land has more than one house superimposed upon it, as in the case of an apartment building, the site assignable to each household is found by dividing the total area of the parcel by the number of households occupying it.

3. CONCEPTUAL FRAMEWORK

We assume that the factors which a household considers in choosing an area in which to locate are its total budget, the items which constitute a market basket, and the costs of obtaining those items. For each household group we posit a set of market baskets among which each household in that group is "indifferent."[13] We posit the set which includes, but is not necessarily limited to, the market baskets currently consumed by households of that type.[14] We permit the household to optimize, not by selecting a market basket from *all* the conceivable sets from which it could obtain satisfaction, but by selecting from the posited set the market basket which maximizes that household's "savings."

These "savings" arise in the following way. A household has a fixed total budget. For a particular market basket the prices of the items in the "other commodities" bundle are given. The residential budget is therefore a residual determined by the size of the total budget and the cost of the "other commodities" bundle. Clearly, it may vary from market basket to market basket. Notice that the character of each of the four items that constitute a residential bundle may vary from market basket to market basket also. Each market basket in the indifference set will have in it a unique residential bundle which has a unique residential budget associated with it. Disregarding site for a moment, the costs of each of the other three items in a residential bundle may vary from area to area. For a particular area, the difference between the residential budget assigned to a particular residential bundle and the cost of the bundle exclusive of the site in it is the maximum amount the household can pay in that area for that site. And it will be the maximum amount that the landowner could extract from the household as site rent. If land were free, it would be a measure of the savings enjoyed by the household because of the locational advantages of the area. We define this difference as the household's rent-paying ability for that site in that area.

---

[13] The household is "indifferent" among the baskets only in a limited sense which will be made clear in the subsequent discussion.

[14] This is based on the assumption that households have, in the past, come close to achieving optimal levels of satisfaction. However, where empirical evidence suggests that market imperfections have precluded optimization we can add to the indifference set market baskets that could be chosen by the household in a market free of imperfections. Obviously there are conceptual weaknesses involved in the use of empirical evidence for the identification of indifference sets. We assume that it is possible to construct operationally acceptable sets that are based on such evidence without being tied rigidly to it. The model will not permit the indifference set that is relevant for a particular household to change during an iteration. But this does not preclude the possibility of allowing taste changes to occur from one iteration to another.

Although we have said that a household is "indifferent" among the market baskets in its indifference set, it seems reasonable to suppose that such savings would have a positive marginal utility for the household. Therefore, in the model, a rational household will attempt to obtain from its indifference set the market basket in which those "savings" are a maximum. In reality, the functioning of the land market may make it possible for the landowner to draw off the "savings" as rent. Nevertheless, the *attempt* of each household to maximize its savings will result in households being allocated to land in configuration that is optimal from the point of view of all the households that are to be located. This allocation will be optimal in a Pareto sense: no household can move to increase its savings without reducing the savings of some other household and simultaneously reducing aggregate savings. Since we have made savings synonymous with rent-paying ability, an optimal allocation is achieved by the maximization of aggregate rent-paying ability.[15]

4. THE PRIMAL PROBLEM

Notation:

U — areas which form an exhaustive subdivision of the region. Areas are indicated by the superscripts $K = 1,2,\ldots,U$.

n — household groups indicated by subscripts $i = 1,2,\ldots,n$

m — residential bundles indicated by subscripts $h = 1,2,\ldots,m$.

b_{ih} — is the residential budget allocated by a household of group i to the purchase of residential bundle h.

c^K_{ih} — is the annual cost to a household of group i of the residential bundle h in area K, exclusive of site cost.

s_{ih} — is the number of acres in the site used by a household of group i if it uses residential bundle h.

L^K — is the number of acres of land available for residential use in area K in a particular iteration of the model.

N_i — is the number of households of group i that are to be located in the region in a particular iteration.

X^K_{ih} — is the number of households of group i using residential bundle h located, by the model, in area K.

---

[15] In a Henry Georgian single-tax economy, the maximization of aggregate rent-paying ability would provide a maximization of public revenue. In a socialist system, if land were free, maximization of aggregate rent-paying ability would provide a maximization of consumer's surplus.

The allocation model:

The primal linear programming model for allocating households to land has the rather simple form:[16]

$$(1.0) \quad \text{Maximize} \quad Z = \sum_{K=1}^{U} \sum_{i=1}^{n} \sum_{h=1}^{m} X_{ih}^{K} (b_{ih} - c_{ih}^{K})$$

subject to:

$$(1.1) \quad \sum_{i=1}^{n} \sum_{h=1}^{m} s_{ih} X_{ih}^{K} \leq L^{K} \qquad (K = 1,2,\ldots,U)$$

$$(1.2) \quad \sum_{K=1}^{U} \sum_{h=1}^{m} X_{ih}^{K} = -N_{i} \qquad (i = 1,2,\ldots,n)$$

$$\text{and all } X_{ih}^{K} \geq 0 \qquad (K = 1,2,\ldots,U) \quad (i = 1,2,\ldots,n) \quad (h = 1,2,\ldots,m)$$

Constraints (1.1) prevent the consumption of land in each area from exceeding the land available. Constraints (1.2) require the model to locate the projected numbers of households of each group. These constraints are equalities because inequalities (of either sense) would not fit the overall requirements of the model. Suppose these constraints were written in such a way that the model was prevented from locating *more than* the projected numbers of households. This would be logical since we are interested in the situation where a particular number of households are located, not where the model can continue locating households in unlimited numbers until all the available land is used up. On the other hand, it is just as logical to write the constraints in such a way that the model is required to locate *at least* the projected numbers of households. This is particularly important where there are household groups which have negative or zero rent-paying ability in all areas. Without constraint, the model would choose not

[16] Readers with limited knowledge of linear programming are referred to Dorfman [4] and the bibliography in Stevens [8].

Although there are *Umn* variables to be determined, it is possible to eliminate many of them in advance of the computation of the program. We can do this for each household in each area by disregarding all residential bundles that yield less than the maximum unit rent-paying ability for that household in that area. These would have to be eliminated in any case in the process of maximization; a prior removal of them can reduce computational time and effort considerably.

to locate these households at all since at best they would not add to, and at worst they would subtract from, aggregate rent-paying ability. For these reasons it is difficult to establish a general rule for the sense of the inequalities. Therefore it is preferable, and perfectly reasonable, to make the constraints equalities.[17] The objective function (1.0) to be maximized is, of course, aggregate rent-paying ability.

Households may be allocated to land in the following ways: (1) One type of household may use all the land available in an area. This will occur where that type of household can yield the highest unit rent for the land in the area and there is a sufficient number of such households to fill the area. (2) The land available in an area may not be used up entirely. Partial utilization will occur where the area has strong locational advantages for only one of the household groups and there are not enough households of this type to fill the area, or where the area has strong locational advantages for two or more household groups and these groups, *in toto*, cannot fill the area. (3) The available land in an area may be left vacant if all households have higher unit rent-paying abilities in other areas and can find sites in other areas. (4) The land available in an area may be used by more than one type of household. This will occur where there are not enough households in the group with the highest unit rent-paying ability in the area to fill that area and they are joined by other households with unit rent-paying abilities the same as, or lower than, the highest group but higher in this area than in other areas. Joint utilization can occur also in the unusual circumstance where two or more household groups have identical unit rent-paying abilities in the area and in all other areas where they could outbid other groups for sites.[18]

---

[17] Actually, it is more likely to be necessary to force the model to locate households with zero or negative rent-paying ability than to restrict the number of households which may be located. This is reflected in the minus signs which appear on both sides of constraints (1.2). If these constraints were written as inequalities they would read: $\sum_{h=1}^{m} \sum_{K=1}^{U} -X_{ih}^{K} \leqslant -N_i$. But without the minus signs the inequalities would be of the opposite sense. In a maximization problem the inequalities *must* be of the "$\leqslant$" form. Therefore the minus signs are necessary. They could be removed when constraints (1.2) are changed to equalities. But the interpretation of the dual variables is somewhat easier if the minus signs are left in the primal.

[18] This is the degenerate case in which there is no unique optimal allocation of the households in the groups which fulfill this condition. A further degenerate case can occur where a particular household group has the same unit rent-paying ability in a number of areas, none of which it can fill completely.

5. THE DUAL OF THE ALLOCATION MODEL

The notation of the dual problem is identical to that of the primal except that the solution variables, X_{ih}^K are replaced by:

r^K the annual rent per unit of land in area K. $(K = 1,2,\ldots,U)$

v_i the annual subsidy per household for all households of group i. $(i = 1,2,\ldots,n)$ The use and meaning of the subsidy variables will be made clear below.

The dual problem is to minimize:

$$(2.0) \quad Z' = \sum_{K=1}^{U} r^K L^K + \sum_{i=1}^{n} v_i(-N_i)$$

subject to:

$$(2.1) \quad s_{ih} r^K - v_i \geq b_{ih} - c_{ih}^K \qquad (K = 1,2,\ldots,U),\ (i = 1,2,\ldots,n),\ (h = 1,2,\ldots,m)$$

$$\text{all} \quad r^K \geq 0 \quad (K = 1,2,\ldots,U).$$

$$v_i \gtreqless 0 \quad (i = 1,2,\ldots,n).^{19}$$

In most linear programming models, the dual presents a problem in interpretation. The existence of the dual is a mathematical fact. But often it also contains information and provides insights which are as important as those provided by the primal itself. This is particularly true in the case of the present model. If we look at the objective function (2.0) and neglect for a moment the second summation, we can interpret the first summation as the total land rent.[20]

It may seem peculiar to minimize total land rent in the dual when we are maximizing aggregate rent-paying ability in the primal problem. It can be shown that the optimal solution of the primal problem must be exactly equal to the optimal solution of the dual. But there is also an important economic interpretation of the dual objective. Suppose that all land in all areas is owned by a monopolist. Then the minimization of site rent will minimize the returns to this monopolist. Alternatively, land could be widely held by individual holders. We would then be minimizing returns to the rentier class as a whole. To put it another way, we are obtaining sites for households as cheaply as possible within the constraints of the model.

[19] An inequality (1.1) in the primal corresponds to a nonnegative variable r^K in the dual. But an equation (1.2) in the primal corresponds to a variable, v_i, whose sign is unrestricted in the dual. Thus the v_i can be positive or negative. See Gale [5].

[20] The second summation is the total of "subsidies" paid to households. We will see later how these subsidies add to the rent-paying ability of households and thereby to the rental income of landowners. But notice that the total value of the subsidies is *subtracted* from the total land rent (and could, in fact, be taxed away from landowners without altering the optimal configuration). Therefore, the value Z' to be minimized is actually *net* land rent.

This is not necessarily a desirable goal if it causes inequities to land owners. But notice that the constraints (2.1) prevent the unit rent on each site from falling below the unit rent-paying ability of *any* household that might locate on that site.[21] This means that the individual landowner can receive at least as much per unit as the highest bidder for his land is willing to pay. This will create certain problems when the household group which can bid highest does not actually purchase the land because it has an even higher unit rent-paying ability elsewhere. It is this latter case, and certain other cases, in which the "subsidy" variables become important.

Bear in mind that a household which can bid the highest unit rent in a particular area is not necessarily of the "wealthiest" household group. Unit rent-paying ability depends upon both total rent-paying ability and size of site purchased. "Poorer" households using small sites may be the highest bidders, per unit of land, in a particular area. Thus "subsidies", in the model, may be assigned in some cases to "wealthy" households.

6. INTERPRETATION OF "SUBSIDIES" AND RENTS

The foregoing serves to introduce the idea that the "subsidy" variables, v_i, are basically a mathematical device.[22] Now let us consider their economic meaning. Suppose, for example, that all household groups save two have been located. Suppose further that all areas have been filled save one. Assume that one household group can bid more per unit for the land than the other. This high-bidding group will then establish the rent level in the area. Because areas are not divisible within the model and because of the nature of linear programming, the same unit rent must then be charged for *all* units of land in the area.[23] The households with the lower rent-paying ability must be located somewhere and this is the only area left. Since their rent-paying ability is insufficient to meet the high rent level established in the area, households of this second group must be "subsidized." And, as indicated above, the same situation would prevail even if the group with the higher rent-paying ability in this area actually locates in another area or areas.

---

[21] Neglecting the v_i, we could divide both sides of (2.1) by s_{ih}. Then unit rent (on the left) must be no less than unit rent-paying ability (on the right). Since this must hold for every household-bundle combination in an area, it then must hold for the combination which would yield the highest unit rent.

[22] For each constraint in the primal problem, there must be a variable in the dual and for each variable in the primal there must be a constraint in the dual. The U land constraints (1.1) are associated with the U variables r^K. The n constraints (1.2) have a similar correspondence with the n variables v_i. The requirement that all households should be located therefore makes the "subsidy" variables a necessary part of the model.

[23] Areas must be fixed in size and number for any run of the model. They may be subdivided or otherwise adjusted after the run.

This simplest interpretation of the subsidy variables is not hard to grasp. However, the existence of these variables creates problems in interpretation which are more difficult to handle. Notice that the v_i are specific to household *groups*. Therefore, once any household in a group receives a subsidy, all of that group must receive the same subsidy. It might be possible to argue that this would be a realistic consequence of an egalitarian public policy, but the argument would be a weak one. For suppose that all "high income" households but one could be accomodated in a particular area. This leftover household would have to locate in another area. If the locational advantages of the first area were extremely large but other areas were highly unsatisfactory to "high-income" households, we might find that this leftover household had a *lower* rent-paying ability in all other areas than competing households of other groups. Since this household must be located somewhere, it will then have to be subsidized. But in the process, all high-income households would be subsidized, raising their rent-paying abilities in the first area to unnecessarily high levels.

We could avoid this problem by disaggregating households and having a constraint (1.2) for each individual household. But this would make our problem extremely unwieldy. A better approach would be to disaggregate areas after a run of the model. This would involve splitting any area in which two or more different household types locate into two or more sub-areas each occupied by a single household type.[24] Rent levels could be adjusted, ostensibly by taxing away from landowners the "extra" rents made possible by the subsidies. Rents would then be equal to rent-paying abilities for all household types and subsidies, of the type we have been discussing, would be eliminated.[25]

Areal disaggregation would enable us to approximate a continuous model in which each small parcel of land is bid on separately and can carry a different rent from neighboring small parcels. The linear programming approximation to the continuous model is especially important in any practical application since the latter is very difficult to solve.[26] Furthermore, the linear programming model may actually be the more realistic of the two. A truly continuous model assumes a level of information and sophistication on the part of both landowners and households which is not likely to exist in practice. From the landowners' point of view it is hard to imagine a real situation

---

[24]Generally, two or more household types will locate in the same area no more times than there are numbers of household groups since the solution to a linear programming model need contain no more non-zero variables than constraints. Our primal problem contains U land constraints and n household group constraints. Therefore no more than $U + n$ variables need be other than zero at an optimal solution. It should be clear that if at least one household locates in every area, no more than n areas can have a second household type.

[25]Subsidies to households with negative rent-paying ability in all areas available to them might still appear, much as subsidies to low-income households may be necessary in the real world.

[26]Cf. Alonso [1].

where there are marked variations in residential rent among contiguous sites; from the households' point of view, it seems unlikely that contiguous sites will be regarded as having distinctly different locational cost and advantages.

We still have not dealt with one important problem. Suppose that a certain household group has the highest rent-paying ability in an area but cannot fill it. Assume that all other households have higher rent-paying abilities in other areas and can be completely accomodated in these areas. This would mean that the land constraint (1.1) for the area would remain an inequality at an optimal solution. In accordance with the general structure of linear programming models, this means that the rent variable corresponding to this area should be equal to zero. In economic theory, land, as a resource for this household type, is then not really scarce and should carry a zero rent. Yet the dual constraints (2.1) for the area require that the land rent be no less than the unit rent-paying ability of any household which might locate there. We appear to have a contradiction unless the households which make the highest bid have zero or negative unit rent-paying ability in the area. If their rent-paying ability is strictly positive we appear to find all units of land in the area earning a positive r^K even though some of the units are not consumed at all. This would seem to make the value of the dual objective function larger than the value of the primal objective function.

We may resolve this conceptual dilemma if we recall that the "subsidy" variables, v_i, need not be positive. Therefore, we can fulfill both the condition of zero rent on unfilled areas and the condition of unit rent greater than or equal to unit rent-paying ability in all areas. The negative "subsidy" can be regarded as a "tax" on the households (equal to the rent-paying ability of the households in the area which they occupy but do not fill) to be used as a "subsidy" to the landowner in the area to "pay" him for the land actually consumed.

By now it may have occurred to the reader that we have obtained the information necessary to construct a classical rent surface for the region. Given a map of the region we could place a block on each area with a height proportional to the unit rent in the area. The surface thus constructed would have discontinuities. But perhaps a discontinuous surface is actually more realistic than the smooth surface of classical theory. A multinodal metropolitan region, with an irregular topography and transportation system and a mixed pattern of land uses, should not be expected to exhibit smooth rent gradients.

The sharpness of the rent discontinuities can be reduced if we can assume that households need not bid the full amount of their rent-paying abilities. If they are able to win land by just outbidding competing households, they can save the difference between rent-paying ability and actual amount bid.[27] The landowner may be able to extract the full rent if, as we suspect, he is the more powerful bargainer.

---

[27] The household could use these savings to purchase additions to any component of its market basket other than site or travel without altering the optimal spatial configuration.

But if there are actually a large number of small competing landowners, their bargaining power may not be much greater than that of the households. Then true net rents may be as high as the rent-paying abilities of the households which occupy the land, as low as slightly above the highest bid of competing households, or somewhere in between, and will depend upon the pattern of land ownership and the vagaries of the land market.[28]

7. SOME LIMITATIONS OF THE MODEL

Many of the limitations of the model are implicit in its construction. They include the following: (1) Data problems: it will often be difficult to obtain consistent data on household budgets and tastes, amenity levels and costs, etc. For the present, in the light of preliminary surveys, we are assuming that it will be possible to obtain data which, though crude, will be adequate to make the model operational. (2) Forecasting problems: successful use of the model depends upon obtaining accurate forecasts of the numbers and characteristics of location-seeking households and of the amounts of available land. (3) Restriction of the choices of a household to a single indifference set: this restriction may prevent the model from achieving a true optimum, since households cannot shift to indifference sets yielding higher levels of satisfaction. Locational savings may or may not constitute an adequate proxy for the satisfactions that might be gained through continuous substitution. (4) Optimization over a number of iterative periods: the model optimizes for individual iterative periods. It may not be rational to assume that an aggregation of such optima will constitute an optimum for the aggregate time period. (5) The cumulation problem: the failure of the model to take into account interactions (other than competition for land) among households locating simultaneously is a serious enough limitation to warrant further discussion.

The problem can be illustrated by considering the cumulation of households around concentrations of job opportunities. Recall that we use a probability interchange model to determine travel cost. This requires the development of a set of expected trip costs for each area on the basis of the number and location of opportunities for trip ends. These expected costs are heavily influenced by the existence of large employment centers which provide trip end opportunities for large numbers of workers.

The difficulty is that too many households may locate in the areas near large employment centers and not enough in the areas surrounding smaller employment centers. At an optimal solution we may find that there are not enough jobs in a large center to accomodate all household wage-earners so that some of them will have to travel relatively long distances to find jobs. This is more a problem with the model than with the real world although there is some evidence to indicate that, given two centers of employment, households tend to orient them-

[28] The discontinuities can be reduced in another way. If we make the areas smaller and smaller (and hence, more and more numerous), the rent surface will be smoother. As indicated above, however, there is both a conceptual and a practical limit to this process.

selves toward the larger one even if they are not sure to find employment there. It may be that near the larger center they expect to have a wider range of choice of employment and be in a better position to find another job should they be laid off.

Nevertheless, the model may give us erroneous results because of the cumulation phenomenon. There are several ways of handling the problem but none of them is completely satisfactory. One approach would be to assign households to jobs in advance of the model run. For example, suppose we were dealing with a metropolitan area with one major and one minor center of employment for workers of a particular household group. Assuming one worker per household, we could then break this group into two parts in proportion to the number of jobs available at the two centers. One part would have its transport costs computed as if only the major center existed. The smaller part would have transport costs relative to the minor center. Such an approach would give strong locational advantages to approximately the right proportions of households near each of the two centers although a successful solution would still not be guaranteed.

An alternative approach would be to make travel costs a function of the number of households locating in particular employment areas. Unfortunately, this would require a non-linear program which is much more difficult to solve. It might provide a theoretical solution to the problem, but not a practical one.

Finally, an approach might be developed out of repeated use of the model. If we find that cumulation occurs, we might be able to determine experimentally how probabilities should be altered or weighted so as to produce a more "realistic" solution.

8. ALTERNATIVE APPROACHES

It is appropriate at this point to discuss the question of alternative models. We have used a linear program because it appears to afford the best compromise between theoretical elegance and practical applicability. A non-linear programming model would be able to handle the cumulation problem as well as problems such as the effect of traffic congestion on travel costs, the effect of growing residential densities on household decisions, and the effects of changing land use on amenity levels and costs. But it will be some time before the new computational techniques for non-linear programming become economically feasible for large-scale problems.

We specifically reject multiple regression models because they tend to project the past into the future in a way which fails to recognize the basic structural interrelationships of activities and land uses.[29] For somewhat the same reason we feel that simulation models do not provide the most promising direction for further research. However, current developments in simulation models for residential decisions indicate that they will give much better results than multiple regression techniques. And it is difficult to ignore the possibility that residential patterns may be more random than rational.[30]

---

[29] E.g. See Chicago Area Transportation Study [3].

[30] Cf. Garrison [6].

Finally, it is important to recognize Alonso's unique contribution in the development of a rigorous theoretical model of residential location. His method is difficult to apply directly, but we feel that our linear programming model provides an analogous approach that is both acceptable and workable.[31]

9. POLICY IMPLICATIONS

The model is applicable to public policy decisions concerning zoning, transportation, redevelopment, public housing, segregation, and other areas of public interest. For example, the model could be run without zoning restrictions. Then restrictions could be applied to determine their effects, if any, on rents and residential patterns. Similarly, the effects of altering transportation costs could be tested by multiple runs of the model.[32]

Where the model gives high rents in areas with little available land and close to existing blighted areas, it may indicate that these areas are ripe for redevelopment. Moreover, the existence of high rent levels would indicate a low cost of writing down the land.[33] Some areas may exhibit extremely low rent levels in comparison with surrounding areas. These may be ripe for redevelopment to raise amenity levels (and reduce amenity costs.) Because the model can identify the "savings" that households might enjoy, not only on vacant land but also in partially improved and built-up areas, it should be of interest in problems of conservation and rehabilitation as well as redevelopment. In all of these applications it could be used as one type of cost and benefit analysis.

We may wish to use the model to consider real subsidies to households which have negative rent-paying ability in all or most areas. Tests could be made to determine whether it is more effective to raise these households' location budgets through direct subsidies (thereby shifting them into another household group) or to reduct their location cost through indirect subsidies (e.g., through public housing). The dual variables, together with knowledge of the total and locational budgets of such households, can provide a measure of how large such subsidies should be.

Notice that the model can recognize racial segregation and similar policies if they have measurable dollar consequences for rent-paying abilities. It can do this in much the same way as it recognizes the amenity value of particular areas. For example, we can ascribe to a predominantly white area high "amenity" costs for particular types of Negro household seeking particular locational bundles.

---

[31] Alonso [1].

[32] Cf. Stevens and Coughlin [9].

[33] And the rent variables may be of use in establishing fair compensation in other types of land condemnation proceedings.

Whenever we use the model as a guide to policy-making, we must be careful that we do not ascribe spurious meanings to its output. There may be social welfare considerations which are non-economic and which are not reflected in the maximization of aggregate rent-paying ability. And these non-economic considerations should be taken into account by the policy maker. He may decide that a true social optimum can be achieved only by deviating from the pattern produced by the model. If he does, the model can tell him how much the deviation will reduce aggregate rent-paying ability. In other words, an "optimal" residential pattern can be found against which to test the effects of alternative programs of public expenditure and control. Although it was designed for use by a transportation study, the model has implications and applications which make it useful in a much broader range of decision problems.

REFERENCES

[1] Alonso, W., *A Model of the Urban Land Market: Location and Densities of Dwellings and Businesses*, unpublished doctoral dissertation, University of Pennsylvania, 1960, microfilmed. A brief description of this model is to be found in Alonso, W., "A General Theory of the Urban Land Market," *Papers and Proceedings of the Regional Science Association*, Vol. VI, 1960.

[2] Carroll, J.D. and H.W. Bevis, "Predicting Local Travel in Urban Regions," *Papers and Proceedings of the Regional Science Association*, Vol. III, 1957, pp. 183-197.

[3] Chicago Area Transportation Study, *Allocating Population To Small Areas*, Chicago, 1957.

[4] Dorfman, R., P.A. Samuelson, and R. Solow, *Linear Programming and Economic Analysis*, McGraw-Hill, New York, 1958.

[5] Gale, D., *The Theory of Linear Economic Models*, McGraw-Hill, New York, 1960.

[6] Garrison, W.L., "Toward a Simulation Model of Urban Growth and Development," paper prepared for the Symposium on Urban Studies, Lund, 1960.

[7] Marble, D.F., "Transport Inputs at Urban Residential Sites," *Papers and Proceedings of the Regional Science Association*, Vol. V, 1959. pp. 253-267.

[8] Stevens, B.H., "A Review of the Literature on Linear Methods and Models for Spatial Analysis," *Journal of the American Institute of Planners*, Vol. XXXVI, No. 3, August, 1960, pp. 253-259.

[9] Stevens, B.H., and R.E. Coughlin, "A Note on Interareal Linear Programming for a Metropolitan Region," *Journal of Regional Science*, Vol. I, No. 2, Spring, 1959, pp. 75-83.

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

THE MEASUREMENT OF INTRA-CITY LOCATIONAL AND RESIDENTIAL PATTERNS

by Beverly Duncan and Otis Dudley Duncan*

Students of urban structure have lived for some time with the uncomfortable realization that their theories--or rather, their abstract, schematic descriptions--of urban growth and form are not very susceptible to empirical testing. Given a map of land uses and residential characteristics, any investigator can discern evidence of concentric patterning, of multiple nucleation, and of sector differentiation.[1] But evidence favoring one of these does not rule out the plausibility of the others, and no way has been found to assess the relative contribution of the three types of tendency to the total configuration of the city. Such an impasse is perhaps typical of the stage of investigation in which natural-history observation has been followed by the formulation of ideal types. Some methodologists believe that it can be escaped only by resorting to more sophisticated measurement devices and formal multiple-variable analysis. The past decade has, indeed, witnessed a good deal of methodological development along these lines. Given the preoccupation of much of this work with what are called "social areas," however, the result has been disappointing. Thus far the advocates of "social area analysis"[2] have accomplished little more than further documentation of the already well-known fact that contemporary American cities are differentiated residentially according to social and economic characteristics.[3] There remains the task of explaining such differentiation and of relating its patterns to basic structural determinants and sequences of growth. Our paper tackles one hitherto neglected phase of this task, exploring some relationships between the pattern of location of industry in the city and the pattern of residential differentiation according to the industrial affiliations of the work force. As a case study of one metropolitan community, Chicago, its purpose is to develop appropriate analytical methods and to suggest hypotheses with some preliminary empirical support.

*The authors are respectively Research Associate at the Population Research and Training Center, University of Chicago, and Professor of Human Ecology, University of Chicago.

[1] The best synthetic discussion of these generalizations is still that of Harris and Ullman [11].

[2] Shevsky and Bell, [20]; Van Ardsol, Camilleri, and Schmid [23].

[3] Hawley and Duncan [13].

Our point of departure is the observation that the economic efficiency associated with elaborate functional specialization of the urban work force requires a more or less pronounced separation of places of work from places of residence and a marked functional differentiation of areas wherein work is carried on.[4] The intra-community pattern of industrial location presumably evolved as part of the total process of city growth, but also in response to such specific locational determinants as space requirements, linkage to transportation and communication facilities, and site characteristics of available land. The pattern of socio-economic residential differentiation likewise reflects the city's history of growth--on this point the concentric-zone and sector hypotheses are agreed--and in particular the several contingencies as to layout of mass-transit routes, residential amenities, and timing of settlement. In our cross-sectional analysis as of about 1950, these locational and residential patterns are taken not as problematic, but as given. Even though we know that they developed together, each changing under the influence of the other, at a given point in time it is possible to characterize them separately and, we believe, to assess at least crudely their interrelation and their independent influences.

Following a description of the data and a presentation of the analytical techniques, the discussion turns to some informal description of the spatial patterns of workplaces and residences. These observations suggest relationships warranting more careful study, making use of a formal analytical model. On the basis of this model it is possible to account for an appreciable part of the areal variation in the proportion of an area's residents affiliated with a given industry in terms of the locational pattern of that industry and the socio-economic level of the area, both factors being independently significant though highly variable in relative importance from industry to industry. Although no satisfactory explanation of this inter-industry variation emerges from our study, its existence argues that the internal residential structure of cities will be found to reflect the industry composition of their economic base and the locational determinants of the basic industries.

1. DATA

The bulk of this research[5] rests on two bodies of small-area data for the Chicago Metropolitan District c. 1950, one pertaining to the distribution of workplaces by industry and the other to the distribution of the resident labor force by industry. Two features of these data call for special comment. First, statistics are available for the entire CMD (Chicago Metropolitan District), an area approximating the City of Chicago plus the contiguous built-up suburban and fringe areas in 1950. The city, suburbs, and fringe together probably constitute a more or less closed labor market. The city or the suburbs and fringe alone clearly do not. Commutation into the city from white-collar dor-

[4] Duncan, B. [4]; Schnore [19].

[5] The initial procurement, processing, and preliminary analysis of the data were accomplished in connection with the Urban Analysis Project of the Chicago Community inventory in 1951-1953. The present analysis was carried out with the aid of a grant to the senior author from the National Science Foundation.

mitory suburbs has received considerable comment, but equally noteworthy are the flows of manufacturing workers from the city into satellite industrial areas. Only fragmentary information on these flows is available, but data derived from the Six-City Labor Mobility Survey show that eight per cent of the manufacturing workers living in the City of Chicago c. 1950 were employed in establishments located in the suburban area.[6] For industries with major suburban industrial concentrations, the proportion of Chicago residents working in suburban areas was more than double the proportion for all manufacturing workers--16 per cent for primary metals, 25 per cent for electrical machinery, and 29 per cent for transportation equipment except motor vehicles. These outflows are countered, of course, by a relatively heavy influx of suburban residents to manufacturing and other establishments located in the City. The areal coverage of the available data, therefore, seems particularly appropriate in that it is co-extensive with the physical and economic city. Second, the areal detail with which manufacturing workplaces are reported and the relatively complete coverage of workplaces are somewhat unique although the data suffer certain serious deficiencies as detailed below.

Workplace data of relatively good quality are available for the two-fifths of the labor force in manufacturing, less detailed and comprehensive statistics are available for the three-tenths in the commercial complex, and no workplace information is available for the remaining three-tenths of the labor force. Manufacturing workplaces by major industry for each of the 935 census tracts comprising the City of Chicago were compiled as a part of the Industrial Land Use Survey conducted during 1950-51.[7] Comparable data for the suburban part of the CMD were compiled from published directories of manufacturers and a mail survey of establishments believed to be large.[8] By comparison with counts of the Census of Manufactures and the Illinois Department of Labor or with the Census of Population count of CMD residents employed in manufacturing industries, these small-area data appear relatively complete. The establishment surveys covered 821,400 workplaces in the CMD, and the 1950 Census of Population reported 828,700 manufacturing workers living in the CMD. The total manufacturing workplace and residence counts differ by only one per cent. Discrepancies between workplace and residence counts are substantially greater, however, for individual manufacturing industries. These discrepancies probably reflect both misclassification of establishments and misreporting of industry by residents, as well as omissions and duplications in the workplace counts. Relative to the labor force statistics reported in the population census on a residence basis, the greatest overcount of workplaces is in the transportation equipment (except motor vehicles) industry, amounting to 28 per cent, while workplaces in printing and allied industries were undercounted by some 30 per cent. Workplace statistics for the other durable goods industry exclude employment in miscellaneous manufactures which was erroneously grouped with unknown industry in the Industrial Survey. (See Col. 2 of Table 1.)

---

[6]The use of these data for measuring the journey to work is described in Duncan, B. [3].

[7]For a description of the survey see Chicago Industrial Study [2].

[8]The compilation of these data is described more fully in Duncan, O. D., and Davis [8].

TABLE 1: Employment in Manufacturing Industries, on the Basis of Census Resident Labor Force Data, in Relation to Number of Workplaces Covered in Surveys of Manufacturing Establishments, Chicago Metropolitan District: c. 1950.

Manufacturing Industry	Per cent of total resident employed labor force	Reported workplaces as ratio to resident employed labor force
All manufacturing	37.8	0.99
(1) Primary metal industries	5.3	1.07
(2) Electrical equipment, machinery, & supplies	4.6	1.01
(3) Other nondurable goods	2.6	0.98
(4) Food and kindred products	4.4	1.00
(5) Other durable goods[a]	2.9	0.58
(6) Printing, publishing and allied industries	3.4	0.70
(7) Chemicals and allied products	1.5	1.07
(8) Apparel and other fabricated textile products	1.6	1.09
(9) Fabricated metal industries	3.6	1.04
(10) Machinery, except electrical	4.4	0.91
(11) Transportation equipment, except motor vehicle	0.8	1.28
(12) Furniture, and lumber and wood products	1.2	1.07
(13) Textile mill products	0.4	0.87
(14) Motor vehicles and motor vehicle equipment	0.8	1.13
Not specified manufacturing industries	0.3	....

[a] Workplace data exclude employment in "miscellaneous manufacturing industries," which are classified under "other durable goods" in the labor force data.

Workplace data available for the commercial complex are less detailed in areal distribution and less comprehensive in coverage than are those for manufacturing industries. Commercial workplace statistics were obtained for the 75 Community Areas making up the City of Chicago and 15 to 25 subareas outside the City.[9] The ratio of workplaces to residences in the CMD was 1.03 for retail and .79 for wholesale trade, .52 for finance, insurance, and real estate, and .74 for hotels and lodging places. Workplace statistics of usable quality were not available for any other industry.

Residential distributions for 14 manufacturing industries and 30-odd other industries were obtained from special tabulations of the census-tract summary cards of the 1950 Census of Population. The industrial composition of workers living in each of the 1,178 census tracts making up the CMD was available.

2. TECHNIQUES OF ANALYSIS

The subsequent analyses rely heavily on a summary measure of the areal distribution of workplaces termed "workplace potential." Workplace potential is analogous to potential of population as defined by Stewart.[10] Instead of dividing persons by their distance away, however, we divide workplaces by their distance away. The workplace potential at a particular site in the CMD is, then, the sum over all workplaces in the CMD of the reciprocals of the distance separating each workplace from the site. We interpret workplace potential as a measure of the accessibility of the site to workplaces in the CMD, on the assumption that accessibility declines as distance increases. Accessibility, conceived as a time-cost factor, need not vary precisely as distance, of course.

Several studies in recent years have demonstrated the utility of population potential and a variant thereof, termed "market potential," each conceived primarily as an indicator of accessibility to the national or a sub-national consuming market, in accounting for inter-city or inter-area differences in economic activity.[11] There has, however, been little application of the potential concept in studies of intra-urban structure and function. Exploratory work with workplace potential measures in the investigation of intra-urban residential structure along the lines reported here was initiated under the Urban Analysis Project, 1951-53.

The computation of potential for a series of sites always involves a number of approximations. Because the locational pattern of industrial activity is more clumped and sporadic than the areal distribution of population, the calculation of workplace potential presents some special problems. The procedure developed for computing potential for

---

[9] Major sources include: *Census of Business* publications of the U. S. Census Bureau; Reiss and Breen [17]; Kitagawa and Sholes [15]; Illinois State Employment Service [14], and unpublished tabulation.

[10] Stewart [22], reprinted in Spengler and Duncan, O. D. [21].

[11] Harris [10]; Duncan, B. [5]; Pappenfort [16].

each manufacturing industry involves first the identification of "major workplace concentrations." Each concentration (minimum of 100 workplaces) on a grid with cells of roughly four square miles was identified, adjacent concentrations sometimes were combined, and workplaces falling outside concentrations were assigned to the nearest concentration. The number of separate concentrations recognized ranged from 11 for textile mill products, an industry with only 7,400 workplaces a fourth of which lie one to two miles from the Loop (CMD central business district) and two-fifths of which lie three to seven miles northwest and west of the Loop, to 54 for the other nondurable goods industry which includes 56,300 workplaces, a sixth located northwest and west some six miles from the Loop, a sixth located southeast some 15 miles from the Loop, and the remainder scattered over the CMD. (Calculations were carried out with 136 concentrations for the "nonelectrical" machinery industry in the exploratory stage.) The method of calculation was analogous to that described in another publication.[12] Potential was computed for a minimum of 29 sites and a maximum of 102 sites, the number depending on the complexity of the configuration of isolines of potential and the judgment of the investigator. For the commercial complex as a whole, workplace potential was calculated using 30 "concentrations," identified less precisely, for 35 sites scattered over the CMD. Workplace potential for an evenly distributed industry was also computed using 191 "workplaces," one located at each intersection of a grid imposed on the CMD.

Isolines of workplace potential for two industries in the CMD that manifest quite different locational patterns appear in Figure 1. The isolines of potential for the fabricated metal industries (A) summarize a relatively even distribution of workplaces over the CMD with no single dominant concentration, although the density of workplaces is comparatively high in the central manufacturing district to the west of the Loop. The area of highest workplace potential is located slightly to the west of the Loop, and potential falls off more or less evenly in all directions as distance from this area increases. By contrast, areas in which workplace potential is high for the primary metal industries (B) occur along the lakefront at distances ten miles or more southeast of the Loop and also at a lesser distance southwest of the Loop. The locational pattern of the industry is shaped by a few large lakefront establishments which together provide half the primary-metals workplaces in the CMD; a smaller concentration of workplaces is found in the central manufacturing district. It should be noted that, although isolines of potential conventionally are shown with equal intervals, in Figure 1 the interval between isolines of potential increases as the magnitude of the potential increases. Had equal intervals between isolines been maintained in Figure 1, the configuration of contours would appear more "peaked," for more closely spaced isolines would be evident in the areas of relatively high potential.

Estimates of industry-specific workplace potentials for 211 quasi-tracts in the CMD were prepared for purposes of regression analysis. The 211 quasi-tracts are a systematic sample drawn from the 1,053 quasi-tracts making up the CMD. (Quasi-tracts are either single census tracts or groups of tracts combined because each had only a few resident work-

[12] Duncan *et al* [6], appendix.

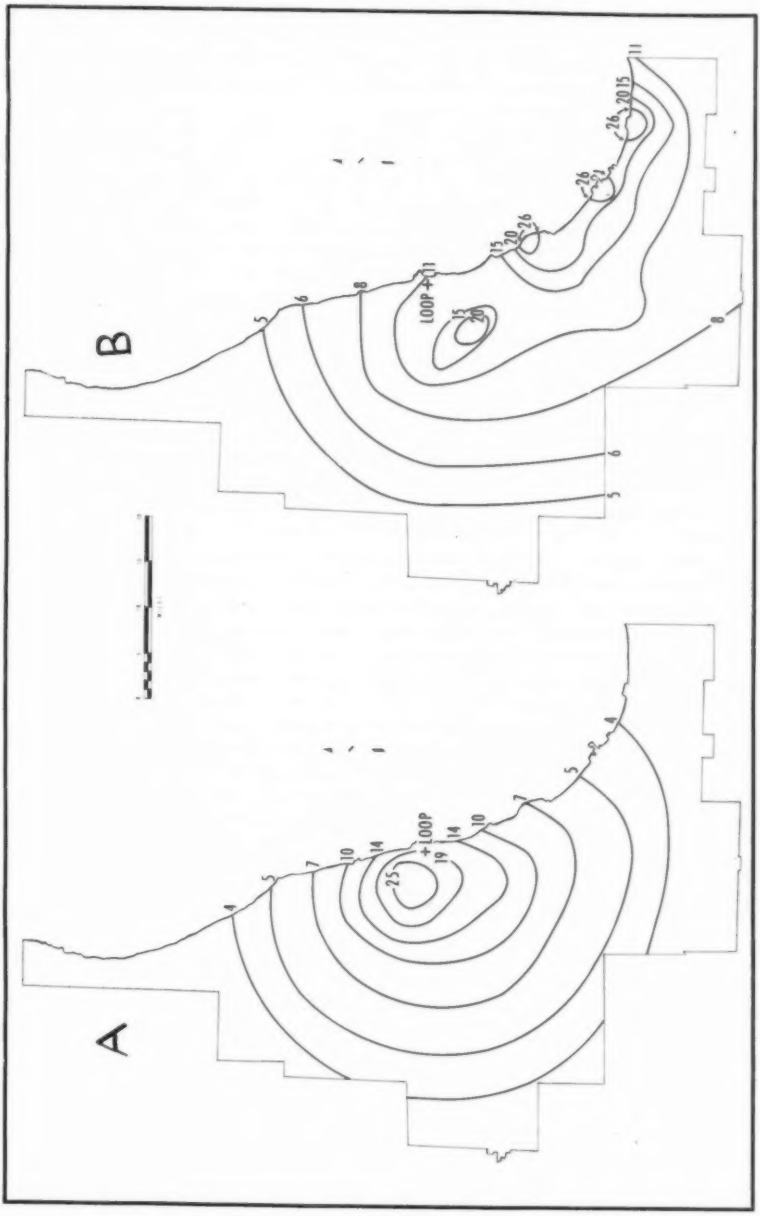


FIGURE 1: Isolines of Workplace Potential, in 1,000 Workplaces per Mile, for Two Manufacturing Industries in the Chicago Metropolitan District: 1951. (A) Fabricated metal industries; (B) Primary metal industries.

ers.) The preliminary estimate of potential for each quasi-tract was obtained by interpolating between isolines on maps similar to but more detailed than those shown in Figure 1. If the quasi-tract contained 250 or more workplaces in any manufacturing industry, the industry's potential in that quasi-tract was computed more precisely by subtracting from the estimate the potential initially attributed to the workplaces within the quasi-tract itself and adding to it a more careful estimate of potential attributable to these workplaces. (These adjustments were required because somewhat crude approximations were involved in the initial potential calculations.) The preliminary estimate of potential for the commercial complex was multiplied by a factor of 1.3 to take account of the fact that workplace coverage was incomplete. This adjustment assumes that the locational pattern of the total complex is identical with that of known commercial workplaces. The preliminary estimate of potential for an even distribution was multiplied by a factor of 3,513 on the assumption that workplaces for the remaining industries--extraction, construction, transportation and utilities, the entertainment and professional services, public administration--are evenly distributed over the CMD. Total workplace potential for each quasi-tract is the sum of 14 industry-specific manufacturing potentials, the commercial-complex potential, and the "other" workplace potential. Hereafter, the proportion of total potential accounted for by an industry is referred to as the industry's "relative potential" as distinct from its workplace potential *per se*.

Another variable playing an important role in the analysis is termed the "expected" proportion of the resident work force in an industry on the basis of occupational composition. A variation on the "method of expected cases" was selected as a technique for assessing the impact of occupation on the residential patterns of industry groups. The 1950 Census of Population provides data on the occupational composition of each census tract in the CMD and a cross-tabulation of occupation by industry for the Chicago Standard Metropolitan Area, 93 per cent of whose resident labor force was living within the CMD. For a particular tract, "expected" residents in an industry is the sum over all occupation groups of the products of the number of residents in a given occupation times the occupation-specific proportion in that industry in the CMD. The "expected" distribution of workers by industry was obtained for each of the 1,178 census tracts in the CMD on a UNIVAC electronic computer. It was then expressed in proportional terms by dividing by the total work force of the area.

If in fact the sole determinant of areal variation in the proportion of resident workers in an industry were occupation, the observed proportion would correspond with the proportion "expected" on the basis of occupation for each tract, and the coefficient b in the equation describing the regression of observed on "expected" proportions would be unity. It has recently been demonstrated that departure of b from unity can be given a meaningful substantive interpretation in analyses of areal variation.[13] When the slope exceeds 1.0, we know that, on the average, the occupation-specific industry proportion in tracts where occupational composition favors the given industry is higher than in the CMD as a whole, and that the occupation-specific proportion tends

[13] Borts [1]; Duncan, O. D., Cuzzort, and Duncan, B. [7].

to be lower than in the CMD in tracts where the occupation composition is generally unfavorable. Or, to put the matter a little less precisely, this kind of result indicates that "factors associated with occupation" in addition to occupation composition *per se* account for part of the variation in industry proportions.

The special merit of the "expected cases" approach for the present purpose is that it allows the "socio-economic status" of the area (insofar as this may be reflected in its occupational composition) to enter the analysis, not in the form of some generalized and vaguely justified "index," but in a form calculated to capture its particular relevance for the dependent variable at hand. The "expected" proportion in a given industry is both an index of the occupational composition of an area, weighted in a particular manner--as may be seen from its definition--and a norm against which one can measure the influence of occupational composition on variation in the industrial attachments of the work force over a set of such areas.

3. PRELIMINARY OBSERVATIONS

Before introducing a model to account for the residential pattern of industry groups, some exploratory work which led to the development of the model should be summarized. A series of studies conducted under the Urban Analysis Project pointed to the fact that industry groups differed from one another both in their locational pattern and in the residential pattern of their work force. Moreover, the workplace and residential distributions did not appear to be independent of one another, for the more centralized residential patterns were those of industries whose workplaces were concentrated toward the Loop. Industries which did not conform, i.e., those evidencing a relatively centralized workplace pattern but a relatively decentralized residential pattern or *vice versa*, typically had an occupational composition quite different from the occupational composition of the population residing in areas near their workplaces. The implication of these results was that the residential pattern of an industry group is shaped by the locational pattern of the industry, the occupational composition of its work force, and the residential patterns of the several occupation groups. Any model which sought to account for the residential patterns of industry groups solely on the basis of the location of industrial activity or solely on the basis of the differentiation of the city by socio-economic level would be demonstrably deficient.

There remained, however, the problems of quantifying the locational pattern of the industry and of indexing the socio-economic character of industries and residential areas. Experimentation indicated that the presence of industrial activity within a residential area was an inadequate measure of that area's accessibility to workplaces. The contrast between a residential area devoid of industrial activity but ringed by industrial districts and a residential area located at considerable distances from industrial concentrations but including a pocket of industrial activity illustrates the point. Applying the concept of workplace potential, which reflects an area's position with respect to the total configuration of industrial activity, each workplace being weighted by the reciprocal of its distance from the area, seems to capture the relevant aspects of the industry's locational pattern. An index of

socio-economic level based on an occupational distribution frequently is constructed by assigning weights to each component occupation and calculating a weighted average. The selection of appropriate weights remains arbitrary, however, in most studies employing such a technique. Our procedure, by contrast, was to weight the occupational distribution, in each tract or residential area by the occupation-specific proportions employed in a particular industry in the CMD as a whole. This means, of course, that the tract has no single index of occupational composition but rather a distinct index for each industry examined. This index, or "expected" proportion of residents in an industry, has special advantages in regression analyses as noted earlier.

Initial studies of the association between industry workplace potential and residential pattern involved a comparison of the residential distribution of workers in a particular industry with that of all workers in the CMD, with the component areas of the CMD ordered in terms of decreasing workplace potential for the industry. The fact that in nearly every case, disproportionate numbers of the industry's work force were living in areas where its workplace potential was high was taken to mean that the locational pattern of an industry shapes the residential pattern of its work force. The three industries evidencing the greatest residential concentration in areas of high workplace potential had a common locational feature: the pattern of each was dominated by a major workplace concentration located some distance away from the districts in which most manufacturing activity occurs. A clustering of small establishments in the Loop area provided two-fifths of the apparel industry employment. Half the primary metals employment was concentrated in a few large establishments along the lakefront southeast of the Loop, the locations of which can be inferred from Figure 1B. A fifth of the employment in electrical machinery was in a single establishment located some six miles west of the Loop. In an attempt to quantify this locational feature, we measured the concentration of the industry's workplaces relative to all other manufacturing workplaces toward the point of peak potential for the industry. The rank correlation of .7 over the 14 industries between the residential concentration in high potential areas and the workplace concentration toward the peak potential suggests that industries whose locational pattern exerts a relatively strong "pull" on the residential pattern are ones with somewhat distinctive and typically mono-centered workplace distributions.

The finding that workers in industries having similar configurations of workplace potential tend to share common residential areas provides further support for the notion that the residential structure is shaped by the locational pattern of industry. A matrix of correlations between the industry-specific workplace potentials for each combination of two of the 14 manufacturing industries over the 211 quasi-tracts was computed. Coefficients of correlation for the 91 pairs of industries ranged from .9 for "nonelectrical" machinery vs. fabricated metals and for the apparel vs. the printing industry to a nonsignificant -.1 for "other durable goods" vs. primary metals. There was a direct and statistically significant correlation between workplace potentials for nearly three-fourths of the pairs. In the corresponding matrix of correlations between the industry-specific proportions in the resident labor force, the coefficients ranged from .6 for "electrical" vs. "nonelectrical" machinery to -.5 for primary metals vs. the printing indus-

try. The correlation between industry-specific residential proportions was direct and statistically significant for nearly a fourth of the combinations and inverse and statistically significant in four cases. The correlation of .6 between corresponding coefficients in the two matrixes (workplace potential and residential distribution) indicates that the residential distributions of two industries tend to resemble one another when their areas of high workplace potential coincide. The industry pairs which fail to conform to this general tendency are, in themselves, instructive. Three of the four cases in which the residential association is unusually low, given the similarity of workplace-potential configurations, involve printing and publishing, a relatively high status manufacturing industry, in combination with some other industry; five of the seven cases in which the residential association is unusually high involve subdivisions of a broader industry group, e.g., "electrical" vs. "nonelectrical" machinery.

These results also suggest something about the variables that should enter a model purporting to account for the residential patterns of industries. The deviant instances just described reaffirm the importance of the occupation factor. The frequent coincidence of areas of high workplace potential for two or more industries suggests that *relative* workplace potentials may be more efficient predictors of residential structure than workplace potentials *per se*. Such an hypothesis would follow if one conceived of total workplace potential as indicative of the total demand for residence in a particular area, on the assumption that demand varies directly with accessibility to workplaces. The industrial composition of the area's resident work force, then, would mirror the industrial composition of its total workplace potential, assuming the effective demand for residence is constant over all industries.

4. THE MODEL

The residential pattern of an industry within the metropolitan district, or more specifically areal variation in the proportion of the resident labor force employed in a particular industry, is the phenomenon to be accounted for. Two probable determinants of residential patterns have been identified: inter-area differences in accessibility to the industry's workplaces; and inter-area differences in socio-economic status (occupational composition). Our model assumes that the proportion of the resident work force in a particular industry is a function of the industry's relative potential, i.e., relative accessibility to its workplaces, and the "expected" proportion of residents in the industry, i.e., socio-economic status. The 14 manufacturing industries are taken to represent 14 "tests" (albeit not wholly independent ones) of the model's goodness of fit.

For each manufacturing industry, the percentage of the total resident work force in the industry varies directly with the industry's relative workplace potential on an area-by-area basis (Col. 1, Table 2). The correlation between the residential proportion and relative potential is significant in the statistical sense in each case, and for half the industries a fourth or more of the inter-area variance in the residential proportion is accounted for by differences among areas in relative potential. Incidentally, the association of the residential proportion with relative potential (Col. 1, Table 2) is stronger than that of the residential proportion with potential *per se* (not shown) for 11 of the 14 manufacturing industries.

TABLE 2: Summary of Regressions of Residential Distributions on Workplace Potentials and "Expected" Residential Distributions, for Manufacturing Industries, Based on Sample of 211 Quasi-Tracts in the Chicago Metropolitan District: c. 1950.

Manufacturing Industry[a]	Correlation and regression coefficients[b]								
	r_{Y1} (1)	b_{Y1} (2)	r_{Y2} (3)	b_{Y2} (4)	r_{12} (5)	$R_{Y.12}$ (6)	$R^2_{Y.12}$ (7)	$b_{Y1.2}$ (8)	$b_{Y2.1}$ (9)
(1) Prim. metal	.86	2.41	.45	2.99	.25	.90	.80	2.23	1.68
(2) Elec. mach.	.65	1.22	.50	2.13	.07‡	.80	.63	1.16	1.95
(3) Oth. nondur.	.78	2.59	.30	2.12	.23	.79	.62	2.49	0.91
(4) Food, etc.	.69	0.74	.41	1.99	.20	.74	.55	0.68	1.38
(5) Oth. durable	.47	0.75	.52	1.49	.12‡	.66	.44	0.65	1.34
(6) Print., pub.	.39	0.55	.36	1.13	-.36	.66	.44	0.85	1.82
(7) Chem., etc.	.64	1.94	.13‡	1.84‡	-.03‡	.65	.43	1.95	2.17
(8) Apparel etc.	.60	1.01	.42	1.23	.36	.64	.40	0.87	0.70
(9) Fab. metal	.27	0.64	.56	1.83	.18	.59	.35	0.41	1.72
(10) Mach., exc. el.	.29	0.80	.54	2.10	.13‡	.58	.34	0.61	1.99
(11) Trans. equip.	.51	1.31	.10‡	1.44‡	.07‡	.51	.26	1.30	0.85‡
(12) Furn., lumber	.39	1.25	.37	2.07	.37	.46	.21	0.94	1.44
(13) Textiles	.37	0.35	.20	0.89	.13‡	.40	.16	0.32	0.68*
(14) Motor veh.	.20	0.23	.32	1.21	.25	.34	.12	0.15*	1.09

[a] See Table 1 for full titles.

[b] Identification of symbols--

Y: Per cent of resident employed persons in the specified industry, by quasi-tract.

X_1: Relative workplace potential of the specified industry, by quasi-tract.

X_2: "Expected" per cent of resident employed persons in the specified industry, by quasi-tract.

r_{Y1}, r_{Y2}, and r_{12}: Zero-order correlations.

b_{Y1} and b_{Y2}: Zero-order regression coefficients.

$b_{Y1.2}$ and $b_{Y2.1}$: Partial regression coefficients.

$R_{Y.12}$: Multiple correlation coefficient, Y on X_1 and X_2.

Coefficients not significant at .05 level are marked (‡); significant at .05 level (*); all other coefficients differ significantly from zero at the .01 level.

It is instructive to look at the regression coefficients as well as the coefficients of correlation. The magnitude of the change in the residential proportion associated with a change in relative potential is substantial for a number of industries. In fact, for half the industries, a change of one percentage point in relative potential is accompanied by a change of more than one percentage point in the residential proportion (Col. 2, Table 2). Inasmuch as the average of relative potential over all areas is very close to the average residential proportion over all areas, the value of the regression coefficient approximates the elasticity of the residential proportion with respect to the relative potential, evaluated at the mean. For the remaining seven industries, a change of one percentage point in relative potential is associated with a change of two- to eight-tenths of a percentage point in the residential proportion.

The percentage of the total resident work force in the industry also varies directly with the area's "expected" percentage in the industry (Col. 3, Table 2). The correlation is statistically significant, however, for only 12 of the 14 industries; and for only four of them does variation in the "expected" proportion account for as much as a quarter of the variation in the observed residential proportion. The "expected" residential proportion or socio-economic level is, on the average, a less efficient predictor of an industry's residential pattern than is relative potential or accessibility to workplaces, as is indicated by a comparison of the correlation coefficients in Columns (3) and (1), Table 2.

The average influence of areal differentials in relative potential on the residential pattern is often less, however, than that of inter-area differences in socio-economic level, as measured by the respective regression coefficients, Columns (2) and (4), Table 2. Possibly a more accurate statement would be that the influence of relative potential on the residential distribution may be substantial or negligible, while the influence of socio-economic level is consistently substantial. A change of one percentage point in the "expected" residential proportion is associated with an average change of more than one percentage point in the observed residential proportion for 13 of the 14 industries; in the other case, it is associated with a change of nine-tenths of a percentage point (Col. 4, Table 2). Moreover, the fact that the values of these regression coefficients exceed unity in nearly all cases means that the residential proportion is higher than "expected" on the basis of occupation in areas where the occupational composition is favorable to residence by members of the industry, and that it is lower than "expected" in areas where the occupational composition is unfavorable. Apparently occupational composition is associated areally with other factors which are themselves correlates of the residential distribution of industries.

Areas in which relative potential for an industry is high need not have occupational compositions which are conducive to residence by members of the industry. They do not in the case of the printing industry, for which an inverse relationship between relative potential and the "expected" proportion in the industry is observed (Col. 5, Table 2). For six additional industries, the "expected" residential proportion varies over areas more or less independently of relative potential. For only half the industries, therefore, do areal differentials in relative po-

tential reinforce socio-economic level in shaping the residential pattern. Most noteworthy is the typically low correlation between the two predictors: in no case is more than a seventh of the variance in "expected" percentage in an industry accounted for by areal differentials in its workplace potential.

As a consequence of this low correlation between the independent variables, accessibility to workplaces and socio-economic level together more fully account for the residential pattern of an industry than does either alone. For four industries, over half the variance in the percentage of the resident work force employed in the industry is accounted for by the combination of relative potential and "expected" residential proportion (Col. 7, Table 2). The two factors combined account for a quarter to a half of the variance in the residential percentages for seven industries and for less than a quarter of the variance for the remaining three industries.

The effect on the residential distribution of each factor with the other "held constant" is significant for all but three industries (Cols. 8 and 9, Table 2). In two of the exceptional cases, only the partial regression coefficient for the residential proportion on relative potential holding constant socio-economic level is significant; in the other case, only the partial regression coefficient for the residential proportion on socio-economic level holding constant relative potential is significant.

Owing to the generally low correlation between the two predictor variables, the pattern of variation by industry of the partial regression coefficients is much like that of the zero-order regressions. Hence the observation stands that occupational composition has a more consistently substantial effect on residential distribution, but that relative workplace potential has the stronger effect for a few industries.

Incidentally, since the variables under examination tend to have somewhat skewed distributions when quasi-tracts are arrayed from low to high, it is well to make sure that our results are not a statistical artifact. Hence some of the calculations were carried out using a square root transformation of the dependent variable. The results were only slightly modified, the correlations being a little higher if anything and manifesting only minor shifts in relative magnitude as among the 14 industries.

5. SOME IMPLICATIONS

We alluded to the fact that industries differ in their residential patterns and then showed that an industry's residential distribution depends in part on its intra-urban locational pattern and the socio-economic differentiation of the city's residential areas. Several aspects of inter-industry variation in residential pattern have been touched on only indirectly, however. The residences of workers in some industries, for example, are distributed over the city in the same pattern as all residences, while the residential distribution for other industries is uneven. We suspect that an uneven residential distribution is evidenced most often by industries whose locational patterns result in substantial areal differentials in accessibility to workplaces and that a coincidence of areas of relatively high accessibility and areas with favorable occu-

pational compositions enhances the unevenness of an industry's residential pattern. In the CMD, nearly half the variance among the 14 manufacturing industries in unevenness of residential distribution (as measured by the ratio of the standard deviation to the mean of the residential proportion) is accounted for by two factors: variability in relative workplace potentials (as measured by the ratio of the standard deviation to the mean of the relative potential) and the coincidence of high potential and favorable socio-economic level areas (as measured by the correlation between relative potential and "expected" residential proportion). This finding provides general support for the notion although the evidence is insufficient as a basis for generalization.

The locational patterns of industries and areal differentials in socio-economic level are taken as given in this study, but they themselves merit attention. Especially important would be an investigation of the way in which the development of the industrial pattern influences the formation of the residential pattern by socio-economic level and vice versa. In the CMD c. 1950, for example, the proportion of professional, managerial, and sales workers among the resident male work force varied inversely with workplace potential for each of the 14 manufacturing industries. (The r's range from -.2 to -.4.) The inverse correlation of the proportion of high status workers with workplace potential, was, in fact, stronger than the direct correlation (r of .2) of socio-economic level and distance from the Loop--the factor that one would emphasize on the basis of extant propositions about areal socio-economic differentiation and city growth.

Inter-industry differences in the efficiency with which the residential pattern is predicted by areal differentials in relative potential and socio-economic level also have been noted. Given the nature of the data, we are certain that errors of measurement obtain in the relative potentials and suspect that the error is not constant over industries. We are reluctant to assume, however, that the observed inter-industry differences are accounted for fully by differential measurement error. The four industries for which the efficiency of prediction is lowest are the smallest of the 14 industries in terms of employment size (see Table 1). Among the ten larger industries, prediction tends to be least efficient for industries whose residential distributions are least variable or most even as compared with the distribution of all residences. Possibly a sizable work force and a variable residential distribution are required if an industry's residential pattern is to be accounted for with gross factors of the type used here.

Industries also differ according to whether inter-area variation in accessibility to workplaces or areal differentials in socio-economic level have the greater influence on their residential distributions. As indexed by the regression coefficients in standard form, the residential pattern is influenced relatively more by areal variation in workplace potential for eight industries and by areal variation in socio-economic level in three cases. For the remaining three industries, the factors have approximately the same influence. The magnitude of the change in the residential proportion per unit change in the independent variable, however, is typically greater for socio-economic level of the area than for its workplace potential, although the reverse is observed for four industries. We are unable to account for the inter-industry differences

in the relative importance of the factors in any systematic fashion, but these differences underscore the observation that areal variation in accessibility to workplaces and socio-economic level jointly and independently shape an industry's residential pattern.

There is reason to doubt that the efficiency of prediction and the relative importance of the two predictors for a particular industry in the CMD would be the same in other cities, although we should expect the residential pattern elsewhere to reflect the locational pattern and the areal differentiation by socio-economic level. The industry "mix"--both the detailed industries included in the major industry groups themselves and the composition of the total work force by major industry--is probably important in shaping an industry's residential distribution. For example, the residential pattern of an industry in a city where that industry was the major source of employment would probably differ from its pattern in a city where it was a negligible source of employment. An industry's residential pattern in a city where employment was concentrated wholly in light manufacturing and nonmanufacturing industries would be unlikely to resemble closely its pattern in a city with a sizable complement of heavy manufacturing.

The issue of how far our findings may be generalized pertains not only to resemblances among cities. A distinct problem is the extent to which findings about the separation of home and workplace for individual workers can be translated into propositions about residential patterns and vice versa. There need be no close correspondence between an industry's rank on the average distance travelled to work by its work force and its rank on a measure of the degree to which residences of its work force are concentrated in areas of high accessibility to its workplaces. In fact, the rank correlation between average distance travelled to work by Chicago residents in each of the 14 manufacturing industries and a measure of the residential concentration in areas of high workplace potential for the same industries is -.35. (The correlation rises to -.6 if the three industries represented in the sample of Chicago workers by fewer than 30 cases are excluded.)[14] Although there is some tendency for industries ranking high on association between residential and workplace distributions to rank low on the average home-work separation of their work force, generalizations based on comparison of the distributions clearly cannot be transferred to home-work separation on an individual basis and *vice versa* with any great confidence. As another example, the average journey to work is substantially longer for Chicago residents working in the Loop than in less central parts of the CMD. However, the manufacturing industry with the highest proportion of its employment concentrated in the Loop--the apparel industry--ranks sixth among the 14 industries on the index of residential concentration in areas of high workplace potential. It requires no examination of data, however, to identify some cogent reasons why the association of the workplace and residential distributions may differ from the association, or average separation, of home and work for individuals in the work force. If there is little inter-area variation in the proportion of the resident work force in an industry, that industry's index of residential concentration must be low. The average journey to work for persons in the industry, however, can be long or short, depending on the locational pattern of work-

[14] These results are based on data described in Duncan, B. [3].

places. Given substantial inter-tract variation in the proportion of residents in a particular industry, the industry's index of residential concentration will be high if areas in which the residential proportion is high are relatively accessible to its workplaces. The average journey to work for its work force, however, can be long or short, depending on whether the areas of high accessibility are proximate to or somewhat removed from workplace concentrations. Industries evidencing considerable variability in their residential proportions will have low indexes of residential concentration should the areas in which the residential proportion is low be areas of relatively high accessibility. The work force of such industries can have a long or short journey to work, however, depending on the locational pattern of the industry. Manifestly, we are confronted here with yet another problem where "ecological correlations" are of truly ecological significance and not mere substitutes for the information needed by analysts of individual behavior.[15]

Despite the restrictions on generalization imposed by having only one city as a case study, the results reported here indicate that any adequate theory of urban residential structure must reckon with both the locational pattern of industrial activity and the socio-economic differentiation of residential areas which comes about through general city growth. To deduce the location of the "zone of workingmen's homes," Burgess, in his schematic presentation of urban residential structure, relied primarily on the latter factor, the sequence of settlement in the course of urban expansion. Characteristics of "workingmen's suburbs," or industrial satellites, have been described,[16] but attempts to reconcile their occurrence with the "zone of workingmen's homes" have been few. Perhaps a more typical response has been the declaration that "ideal type zones" do not exist. We believe there is more promise in a model that assumes both areal differentiation in socio-economic level, stemming at least in part from the process of city growth, and areal differentiation in industrial activity, based on determinants of industrial location.

REFERENCES

[1] Borts, G.H., "Comment (on 'Analysis of Interstate Income Differentials,' by Frank A. Hanna)," in Conference on Research in Income and Wealth, *Regional Income*, Studies in Income and Wealth, Vol. 21 (Princeton: Princeton University Press, 1957).

[2] Chicago Industrial Study: *Summary Report* (Chicago: Chicago Plan Commission, 1952).

[3] Duncan, Beverly, "Factors in Work-Residence Separation: Wage and Salary Workers, Chicago, 1951," *American Sociological Review*, 21 (February, 1956), pp. 48-56.

[4] Duncan, Beverly, "Intra-Urban Population Movement," in Hatt and Reiss, [12].

[5] Duncan, Beverly, "Population Distribution and Manufacturing Activity: The Nonmetropolitan United States in 1950," *Papers and Proceedings of the Regional Science Association*, 5 (1959), pp. 95-104.

[15] Cf. the statement of this point by Duncan, O. D. and Lieberson [9].

[16] Schnore [18].

[6] Duncan, Otis Dudley, *et al.*, *Metropolis and Region* (Baltimore: Johns Hopkins Press, 1960).

[7] Duncan, O. D., R. P. Cuzzort and B. Duncan, *Statistical Geography* (Glencoe, Ill.: The Free Press, 1960), section 3.4.

[8] Duncan, O. D., and B. Davis, *Ecological Aspects of the Labor Force in the Chicago Metropolitan Area*, Urban Analysis Project Report No. 20 (Chicago: Chicago Community Inventory, University of Chicago, hectographed, 1953).

[9] Duncan, O. D., and S. Lieberson, "Fthnic Segregation and Assimilation," *American Journal of Sociology*, 64 (January, 1959), pp. 364-74.

[10] Harris, C. D., "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 44 (December, 1954), pp. 315-48.

[11] Harris, C.D., and E.L. Ullman, "The Nature of Cities," first published in 1945 and reprinted (among other places) in Hatt and Reiss [12].

[12] Hatt, P. K. and A. J. Reiss, Jr. *Cities and Society*, (Glencoe, Ill.: The Free Press, 1957).

[13] Hawley, A. H., and O. D. Duncan, "Social Area Analysis: A Critical Appraisal," *Land Economics*, 33 (November, 1957), pp. 337-45.

[14] Illinois State Employment Service; Division of Unemployment Compensation, *Analysis of Employment by Place of Work and Place of Residence: Chicago Area (Cook and DuPage Counties)* (mimeographed, 1949).

[15] Kitagawa, E. M., and D. Sholes, *Chicagoland's Retail Market* (Chicago: Chicago Association of Commerce and Industry and Chicago Community Inventory, University of Chicago, 1957).

[16] Pappenfort, D. M., "The Ecological Field and the Metropolitan Community: Manufacturing and Management," *American Journal of Sociology*, 64 (January, 1959), pp. 380-85.

[17] Reiss, A. J., Jr. and L. Z. Breen, *Geographic Distribution of Retail Trade in the Chicago Metropolitan Area, 1948* (Chicago: Community Inventory, University of Chicago, 1952).

[18] Schnore, L. F., "Satellites and Suburbs," *Social Forces*, 36 (December, 1957), pp. 121-27.

[19] Schnore, L. F., "The Separation of Home and Work: A Problem for Human Ecology," *Social Forces*, 32 (May, 1954), pp. 336-43.

[20] Shevky, E. and W. Bell, *Social Area Analysis*, "Stanford Sociological Series," No. 1 (Stanford, Calif.: Stanford University Press, 1955).

[21] Spengler, J. J. and O. D. Duncan, editors, *Demographic Analysis: Selected Readings* (Glencoe, Ill.: The Free Press, 1956).

[22] Stewart, J. Q., "Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population," *Geographical Review*, 37 (July, 1947), 461-85.

[23] Van Arsdol, M. D., Jr., S. F. Camilleri, and C. F. Schmid, "The Generality of Urban Social Area Indexes," *American Sociological Review*, 23 (June, 1958), pp. 277-84.

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

THE METROPOLITAN CITY REGION OF THE MIDDLE AGES

by Josiah C. Russell*

The existence of a pattern of cities by size within a region is well known to students of modern society.[1] A study of medieval population also reveals significant relationships among the larger cities of that time.[2] The medieval age was a much simpler time and therefore, in theory at least, the organization of its culture should have been less complicated and easier to understand. On the other hand we do not have as accurate information about medieval city sizes and patterns as we have about modern cities.

Modern large city regions are often delimited by such things as the circulation of metropolitan newspapers, the relationships between central and outlying banks, and the market and supply areas of wholesale and industrial concerns.[3] Not all of these phenomena delimit the same territory exactly; thus the limits of the region are often indefinite. Furthermore, regions seldom, at least in the United States, coincide with political boundaries, and so do not necessarily have an integrated political life. Thus the region is not as well defined as the state or nation. Similarly, the medieval region was not as well defined as kingdom, duchy, city-state or ecclesiastical province unless it happened to coincide with one or more of them.

Medieval students, indeed, would have found the topic of 'the region' an ideal one for scholastic discussion. The Nominalists would have said that 'regio' was merely a convenient term for a large area with perhaps some uniform characteristics surrounding a great city. The Realists would have insisted that the 'regio' had a definite type of life which differentiated it from other areas. Although each side

---

*The author is Professor of History at the University of New Mexico.

[1]See Berry and Garrison [3], and literature cited there. Also, Christaller [7], Lösch [18], Philbrick [20], Ullman [34], and Dickinson [8].

[2]Russell [23], pp. 68-71.

[3]Lösch [18], pp. 366, 404.

could have marshalled impressive arguments to defend its position, the point of view of the realists would have been the easier to understand since medieval society was visually minded in an age when one could see and understand nearly every phase of life.

Since medieval regionalism is a large and intricate subject, a study such as this cannot be very detailed.[4] It is designed to point to possibilities for further study and to emphasize the significance of the subject. The study is limited to a particular type of region: the area associated with the largest city of a district of about 30,000-100,000 square miles. First, the types of information used to delimit the region will be presented: for example physical, political, religious, economic and intellectual characteristics. Then the distribution of population will be considered. Third, the effects of distance and transportation in determining the location of settlements will be taken up. It is hoped that the study will indicate the range and importance of the metropolitan city region in medieval society.

1. THE CHARACTERISTICS OF THE MEDIEVAL REGION

The most obvious way to delimit the region is by political or ecclesiastical boundaries; both political administration and ecclesiastical discipline operated in themselves to produce a common life. And yet accident of conquest or political marriage sometimes placed in the same jurisdiction areas which were really not parts of the same region. A region might have had an intellectual life based upon the same language, schools, dialect or even (in the later Middle Ages) the same universities. While the social life of the various classes of feudal society usually coincided with their political life, at times it was distinct. The economic limits of the region are often more difficult to define, although they are probably the most important. Patterns of migration and population distribution sometimes afford a very good means of regional delimitation.

A tentative pattern of location of metropolitan cities in the thirteenth century (Figure 1) is offered as an illustration of this latter approach. Much uncertainty exists about the size of many of these cities, and this makes it difficult to determine their relative status within regions. But, there does appear to be some spatial regularity.

Another way of delimiting regions is in accordance with the configuration of the land. Large cities and their surrounding areas are often set off by sea or desert. Examples of these are Cairo and Egypt, Tunis and northwest Africa, Palermo and Sicily and London and England. Then there are the great peninsulas containing more than one metropolitan city region: Italy (Rome, Naples, Milan), Spain (Cordoba, Barcelona, etc.), Greece and the Balkans (Constantinople, Saloniki, Athens, Smyrna, etc.).[5] Even here it is not easy to find the limits

[4] Probably the medieval German cities have been most carefully studied: See Dickinson [9], pp. 9-22, 47-53 and bibliography: and [10].

[5] In this area the widely varying size of Constantinople created a question as to the regional relationships. See footnote 26.

FIGURE 1: European Cities in the 13th Century.

exactly. It is harder yet to delineate the regions which depended upon the great continental cities, such as Paris, Cologne, Budapest, and Kiev. The evidence which we must seek in order to delineate such regions is the existence of a network of settlements of various sizes forming a definite pattern about the major city.

In many cases the physical limits of a region have been also those of a political unit. This has been particularly true of such well defined regions as England, Egypt, and Sicily. The boundaries on the map which set off the area as a region easy to organize from within, were also easy boundaries to defend. These considerations of organization and defence naturally tended to create a common feeling in the area. The late medieval kingdom of Aragon-Catalonia was apparently a region, as was the Kingdom of Naples in the same period. On the other hand the scattered holdings of the French kings in most of the Middle Ages could hardly be identified as a region, and the Holy Roman Empire included more than one region.

In a religiously inclined age the ecclesiastical organization is always important. Christianity spread initially among urban groups, so that its chief officials, the metropolitans, patriarchs and archbishops, were leaders in the largest cities: Rome, Alexandria, Antioch, Ephesus, Carthage and Lyons among others. The decline of the Roman

Empire threw effective civic responsibility onto the bishops.[6] The rural churches were subordinate to the city churches. Even the Archbishop of Canterbury usually stayed at Lambeth Palace just across the Thames from Westminster rather than in the small city of Canterbury. An exception was Paris which had only a bishop who was subordinate to to the archbishop of Reims. This may help explain why Paris, even though capital of Capetian France, grew so slowly in the twelfth and thirteenth centuries. The province of Cologne may possibly have constituted a region, although the city had strong connections in the upper Rhine as well as in the lower valley. In many cases bishoprics and even provinces (the archbishop's jurisdiction) were so small that they were just divisions of regions.

The large city center or metropolis was frequently the political capital of the region. This was true for London in later medieval England, for Naples in southern Italy, and for Kiev in the southern part of medieval Russia. Milan was a political center for northern Italy for long periods. Paris, politically the French capital, seems to have been merely a regional center of northern France until near the end of the Middle Ages. In general, the capital enjoyed the protection of the sovereign and the wealth of the ruling classes. It is not surprising that the political function contributed to the growth of metropolitan cities.

The economic functions of the region also tended to center in the metropolis which was often the hub of a road system inherited from the Romans or was conveniently located for water transportation. The extent of financial operations is not as easy to define as the political boundaries in medieval life. In fact, localism rather than regionalism, has been emphasized as the chief characteristic of the economic life of the time. Yet the thirty-fifth article of Magna Carta (1215) prescribes standard measures for wine, beer, grain, cloth and even weights for the whole kingdom of England.[7] Here was a strong effort to enforce a regional uniformity in line with the practice of the metropolitan city. Common currency and great trade fairs were important factors in the regionalization of the economy. The region was nearly always an economic unit whatever else it might have been.

Two recent publications have shown that migrants into the cities retained connections and even land holdings in their native villages and cities.[8] The medieval custom of taking the name of one's home town as a surname after migration enables the student to follow the course of migration rather more easily than might be imagined. The intermingling of persons within the region tended also to produce a regional physical type or types as some married others who conformed to the physical type and the atypical moved into the cities.[9]

---

[6] A condition legalized by Justinian's Pragmatic Sanction of 554. See Villari [36], pp. 253-254.

[7] See Stubbs [32], p. 297.

[8] Ekwall [11], Russell [24], Wolff [37].

[9] Russell [25].

The region often had distinguishing intellectual characteristics. Southern France was the land of the written law, (i.e., the area of the survival of Roman Law), whereas northern France was the land of the customary law.[10] In southern France (Figure 2) there was a distinct form of Romance Language which caused that area, of which Toulouse was the metropolitan city, to be called Languedoc.[11] In its eastern part, some of the Rhône Valley was probably a part of the region dominated by Lyons, at least after the Burgundians established a kingdom there.

In a visually minded age the more obvious features of life were apt to be common throughout a region or a subregion. Dress, architecture, wines and cheese, among other items, developed regional characteristics. These give variety to parts of Europe even today and gave even more before the Europeanization and Americanization produced by the Industrial Revolution. Changes of regional characteristics did occur, e.g., the wholesale shift from early English to Norman names about 1120-40 in England. This development of a 'consciousness of kind' unfortunately resulted in an antagonism for alien elements, notably the Jews. Indeed, medieval regionalism sometimes presaged modern nationalism but lacked its main feature, the 'desire to live together under one government.'

The region was thus an expression of the folkways of its inhabitants. In addition it developed a demographic structure in terms of distribution of population.

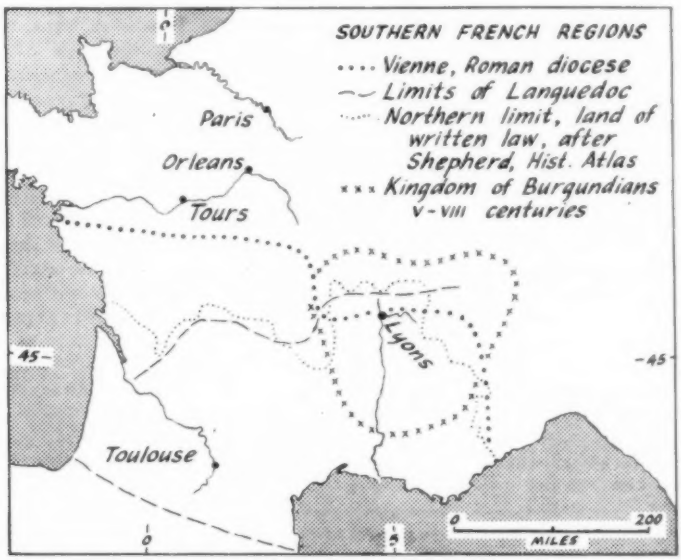


FIGURE 2: Regions of Southern France.

[10] For its limits see Shepherd [27], p. 147.

[11] For one map see Holmes and Schutz [14], p. 44.

2. THE DISTRIBUTION OF THE POPULATION

The study of population traditionally has been directed toward numbers of people, birth, death, and marriage rates, and age distribution. The existence of patterns of settlements and of the causes of these patterns has seldom been considered. Today, much attention is given to hypothetical models of population distribution and classification of cities. Hypothetical models usually are based on simplified assumptions which are not realized in practical application.[12] Thus, actual patterns should be carefully studied to test the models. The earth seldom presents relatively uniform conditions over great areas, except for such uninhabitable environments as ocean, ice, and desert. Nevertheless man does seem to aim for economy of effort,[13] and the result is an ideal pattern modified by physical conditions.

Sometimes the Middle Ages is considered as an age of localism, with little commerce. If this were true we would expect that no large commercial cities would arise.[14] However, this was not so. There were always some cities, although the small total population in much of the European-Mediterranean territory naturally kept the size of the cities small. Great trading cities such as Constantinople, Antioch and Alexandria existed in the east, while in the west Cordoba, Toulouse, Cologne, Milan and others were of moderate size. The problem is to find the amount of urbanization and the pattern of distribution.

Not many areas provide data adequate to determine the percentage of urban population. In England, data from the Domesday Book of 1086 and the poll tax of 1377 offer some information. At the time of Domesday the boroughs seem to have had about 10 per cent of the population.[15] The boroughs taxed separately in 1377 had about 114,266 taxpayers out of a total of about 1,417,000 taxpayers.[16] However, other places called boroughs (at least 97) were included with the county returns, so that the true urban total was again about ten percent.

A study of the empirical regularities in population distribution suggests the existence of a 'physics of population distribution.' In other words, masses of population appear to behave in a manner analogous to the behavior of physical masses.[17] It has also been suggested that population tends to adjust to a most economical pattern. This

---

[12] For an ideal network see Dickinson [8], p. 30, a modern case but theoretically applicable to medieval conditions.

[13] For a general approach see Zipf [38]. The main principle is defined (p. 6) as the 'least average rate of probable work.'

[14] A dispute over the existence of cities before the Norman Conquest has on the affirmative Tait [33], and on the negative Stephenson [29].

[15] Russell [22], note 2, p. 54.

[16] Russell [22], p. 146.

[17] Stewart and Warntz [31] esp. pp. 102-106; Stewart [30].

pattern has been postulated to be hexagonal.[18] Modern instances show that settlements do perhaps tend to form a hexagonal pattern about cities. The overlapping of different sized hexagonal patterns can be made to produce 12 alternating sectors of dense and sparse settlement around a major city. These sectors gradually merge into a relatively even distribution of agricultural settlement with increasing distance from the city.

This can be tested, for England in the eleventh century, by information available in the Domesday Book,[19] which gives number of holdings by manor or village for most of England. As an example, Leicestershire is used because it is an uncomplicated area with its one city, Leicester, near its center (Figure 3). The villages do seem to fall roughly into a six spoke pattern. However, this pattern is obviously modified by the terrain. Charnwood Forest provides an extra

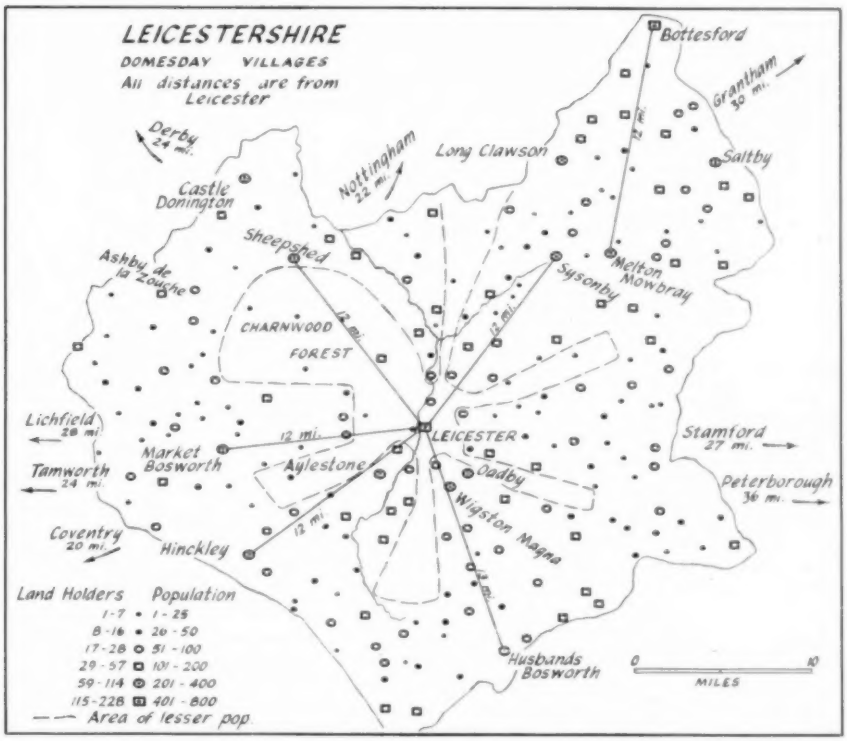


FIGURE 3: The Settlement of Leicestershire in 1086.

[18] Lösch [18], p. 125.

[19] [35] pp. 277-354, map opposite p. 306. I have added together the numbers for each place as they are occasionally scattered.

wide break between spokes in the northwest, while valleys determine the shape of the central and eastern spokes. The total number of places (about 300) divided into Leicestershire's 824 square miles produces an average of about 2.75 square miles per place, but if Charnwood Forest is omitted, the average is reduced to about 2.58 square miles. This would place the villages on the average about 1.8 miles apart, the towns of population 100 and more (about 55) about 4.2 miles apart, and those settlements of more than 200 about 8.8 miles apart. The average distance from Leicester to small boroughs is about 13 miles, while the distance to the nearest places of equal size is 20 to 27 miles. This fits the hypothetical pattern of population distribution quite well.

Theoretically, the pattern would show the largest city at the center of the region with six satellite cities of about equal size and at an equal but considerable distance from the metropolitan city. Actually, a highly skewed pattern for the satellite cities nearly always appears. This is probably because the largest city is seldom in the center, and the distribution of resources is usually irregular. However, the distribution of medieval cities by size does seem to conform to the pattern suggested by this formula:[20]

$$n_r = \frac{C\left(1 + \frac{\sqrt{r-1}}{10}\right)}{r}$$

r = rank of city by size
n_r = population of city of r rank in region
C = size of the largest city

The result is that the cities should show the following proportionate size: 100, 55, 38, 29, 24, etc. If the largest city has 1.5% of the total regional population, the largest hundred places should have about ten per cent of the total population. Or to put it the other way, if the cities (assuming about 100 of them) equal ten per cent of the total population, the largest city would have about 1.5 per cent.

The convergence of the rank-size series is indicated by the fact that the total population of the sixty-five largest cities is about nine per cent of the regional while for the 145 largest it is about twelve per cent. Obviously the smaller of the 145 largest settlements within a region would be merely market towns. So we may assume that about 10% of the population can be found in settlements which can properly be called cities. It might be added that the rank-size formula apparently does not hold for settlements much below the size of cities as we understand them; below a certain size, the villages seem to conform to the law of chance.[21] Conditions in the various medieval regions

[20] Used in Russell [23], p. 69. This is a slight modification of Zipf's formula.

[21] Russell [22], pp. 307-12.

were sufficiently similar that the pattern is quite consistent, although there are exceptions. One exception occurs when two or more closely situated cities seem to take the place of a single city in the pattern. The hypothetical pattern implies that cities of the first rank will be at a considerable distance from each other. Failure to be at the proper distance apart would indicate that the cities share in the functions of their part of the region.[22]

The distribution of cities in England in 1377 (Figure 4) seems to follow the above pattern. London had about 1.5% of the total population, about 35,000 inhabitants. The second rank city, York, had 11,000 population instead of the expected 20,000. But the addition of its port, Hull, and Beverly, another city near Hull, makes a group of

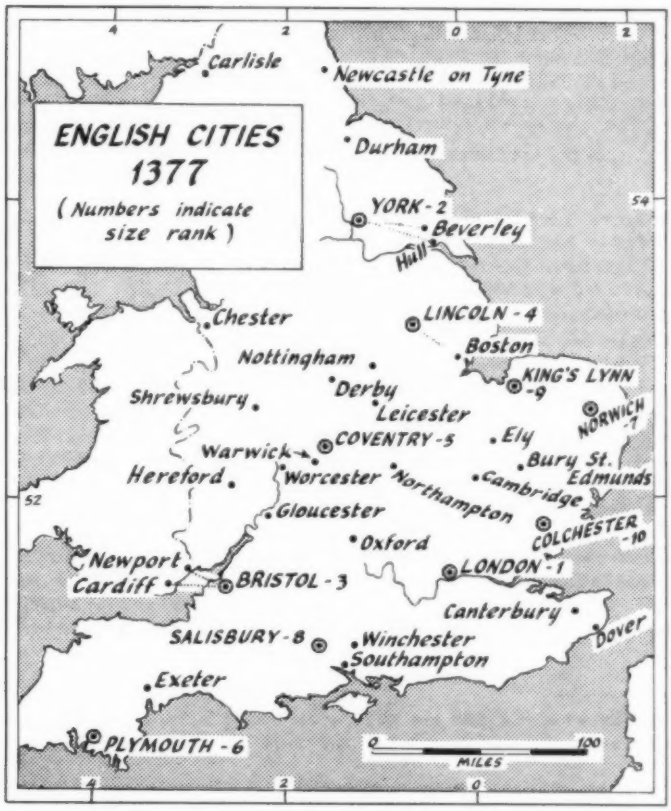


FIGURE 4: English Cities, 1377.

[22]This may be true of the Bremen-Hamburg-Lubeck area of North Germany.

about 17,000.[23] The third rank city, Bristol, had only about 9,500, but Cardiff and Newport with several thousand more lay just across the Bristol Channel about twenty and fifteen miles away respectively. Lincoln (5,350) and its port, Boston (4,300) on the Wash seem to constitute a fourth group.[24] Within the groups the cities are all connected by navigable water (except possibly York and Beverley), which makes their close association likely. Coventry (7,200) together with Warwick,[25] Plymouth (7,200), Norwich (5,900), Salisbury (4,800), King's Lynn (4,700) and Colchester (4,400) follow in approximately the right size order. The accumulation of major economic and political power in the larger cities other than London was probably prevented by the efficient administration in the counties by the powerful central government.

The map shows the dispersal of the larger cities away from London. Perhaps the large size of Plymouth, so far to the west and away from the greater centers of population, is surprising. However, it lay close to the mineral wealth of Cornwall and Devon and had long been a shipping center. All of the ten largest cities, except Coventry, were near the sea or on a navigable river. Perhaps Coventry was somewhat smaller than it might have been because of the costliness of land transport.

There are instances in the medieval world where regional metropolitan cities had populations in excess of 1.5% of regional populations. Alexandria must have had a much larger percentage of Egypt's population: it was the economic center of the whole eastern end of the Mediterranean. In the Low Countries in the later Middle Ages the great textile centers, such as Ghent, were far larger than 1.5% of any region of which they were centers. Similarly, Genoa and Venice were centers of maritime rather than land regions. On the other hand the Carolingian Empire, even at its height about 800 A.D., had no city with even one per cent of its several millions of population. This may be evidence of the primitive quality of both the economic and political life of the Empire. These, however, were exceptions to the general uniformity of regional patterns.

The effect of changing function upon city size is evident in the case of Constantinople. As the rather unimportant city of Byzantium it had, at the beginning of the fourth century, perhaps 20,000 inhabitants.[26] Constantine made it the administrative center of the

---

[23] Beverly with almost 4,000 and Hull with about 2,350. The whole Humber-Ouse Valley was a center of trade. For York see Bartlett [1].

[24] Nearby was also a profitable woolen industry. See Carus-Wilson [6]. Lincoln and Boston are about thirty miles apart on the Witham.

[25] Warwick's population is not known but it must have been several thousand, they were only ten miles apart.

[26] For its population changes see Russell [23], pp. 66, 77, 81, 93, 99, 129-30.

eastern half of the Empire, the New Rome of the East. In his time its population rose to more than 80,000 and passed the 150,000 mark before another century had gone by. At the end of the Middle Ages it declined as did the Empire itself. Just before its capture by the Turks in 1453, Constantinople was the capital of a much reduced Empire and probably had a population of about 35,000. The Turks, calling it Istanbul, made it the capital of another great empire which again brought the population of the city well above 100,000.

3. TRANSPORTATION AND DISTANCE

Although medieval cities were not large by modern standards and their inhabitants comprised only a small proportion of the total population, they still could not supply themselves exclusively from their immediate neighborhoods. Commerce and trade were necessary and involved transportation of various kinds. The non-economic aspects of regional life also required communications and transportation. The friction of space which these had to offset was an important factor in the regular pattern already outlined. Just as they are today, transportation and distance were very important factors in regional activity in the Middle Ages.[27]

The cost of land transportation in medieval times has been variously estimated.[28] In Diocletian's time the official price of a cartload of hay was 600 pence (*denarii*), but its transport cost was, at a maximum, about twenty pence a mile. This made the transport cost of a cartload equal its value in thirty miles.[29] In thirteenth and fourteenth century England land carriage was estimated at about a penny per ton-mile for wheat.[30] At the price of five shillings a quarter (eight bushels) the ton would be worth about 21 shillings or 252 pence. Thus the cost of transporting wheat for twenty-five miles would be about one-tenth of its value, still a high price. The carriage of 125 pounds of salt (about five pence worth) for eighty miles was five pence.[31] The transport rate usually jumped sharply for hauls taking more than one day. For example from Southhampton to Oxford the carriage charge on a wagon load of wine was 3.5 pence a mile in 1395.[32] In contrast, by the end of the twelfth century a remarkable land transport system had developed between the Mediterranean and the Fairs of

---

[27] Lösch [18], pp. 182-191, 427-30.

[28] Lopez [17]; Rogers [21], I, pp. 661 and 656-663.

[29] Lopez [17], p. 18.

[30] Rogers [21], I, p. 658.

[31] Rogers [21], I, p. 457. For the period after 1400 see Rogers [21], IV, pp. 692-713.

[32] Rogers [21], I, p. 660.

Champagne (and Flanders).[33] This system was able to compete with the water transportation of the time, despite the fact that water transportation was generally much cheaper.[34]

Even within the village,[35] distance was a significant factor since carriage was a problem, and one usually walked to work.[36] Within the limits of the village proper were gardens and often stock pens. The dwellings were surrounded by the cultivated land and beyond lay meadows, orchards, pasture and woods.[37] Nearby land was cultivated more intensively since a walk of more than a mile to such hard work as plowing, harrowing or harvesting was a heavy burden.[38] Even for lighter work the walk was a chore and tended to limit the manor to a few square miles. The number of manors in medieval England has been estimated at about 9,000. Given the total area of England, the average manor would consist of about 4.5 square miles.[39]

However, the problem of distance to work merely defined an outer limit to the manor. The actual area covered by a village often was much smaller than the potential area and might have been the result of folk ideas of settlement. A Celtic hamlet would have an organization quite different from that of a Germanic village. In certain areas of England, villages are spaced consistently between 0.9 and 1.2 miles apart: the regularity indicates a definite folk pattern.[40] The curious Jutish settlements in Kent were apparently carried over from the Rhineland.[41]

Different types of economy produced different types of settlement. Mediterranean lands, relying heavily upon olives and grapes had forms of settlement differing from those of the colder and moister northern lands. A pastoral economy, such as that of Spain and North Africa, would have larger units than one relying more upon arable land.

---

[33] Face [12].

[34] Face [12], p. 246.

[35] The village was the smallest settlement in the Middle Ages since only rarely did people live in isolated farmhouses or in very small hamlets.

[36] "And he (the bailiff) ought to forbid any provost or bedel or hayward or any other servant of the manor to ride on, or lend, or ill-treat the carthorses and others." Lamond [16], p. 91.

[37] Beresford and St. Joseph [2], pp. 47, 256.

[38] Lösch [18], p. 381 citing Müller-Wille [19], a modern situation apparently coming from the Middle Ages.

[39] Larger manors often were divided into parts, sometimes called berwicks or wicks. An example of modern spacing deriving from the Middle Ages is in Bracey [4].

[40] Beresford and St. Joseph [2], pp. 75-76.

[41] Jolliffe [15].

Nor should one forget that all settlements were set up under the jurisdiction of a ruling class by its permission and under its direction. Great areas were reserved by feudal lords for hunting and thus only gradually came into use for agriculture. The forest was a very real limiting factor.[42]

Since grain and other heavy agricultural products made up the bulk of transported commodities, we may classify settlements as: (a) those producing a surplus and thus feeding other villages as well as themselves; (b) those which had just sufficient for themselves; and (c) those importing from others because they were not self-sufficient. In general, if a manor or village included about 2 + square miles of good farmland it ought to have had at least a square mile of arable land (640 acres). At twenty acres to the family this meant 32 families and a population of about 140.[43] The manors up to this size, being almost completely agrarian, should have exported grain. The medium-sized manors (from 200 to 400 people)[44] probably supported substantial numbers of non-agricultural workers and had little grain to export. Larger towns had to import grain. Their modest need for imports generally was satisifed with a weekly market. If the market towns were 6-8 miles apart,[45] they would be close enough that peasants from the smaller places could walk to and from the markets.

For the small city, the day's journey seems to have been the ordinary limit for bringing in food and some other necessities. This should have been about thirty miles at the most, but for practical reasons it was often less. Frequently, the farmer would bring in his products after the day's work. The produce kept better in the cool of the night and the farmer could sleep while travelling and arrive in the market place early in the morning. He could finish his business, selling his produce and buying for his own needs, by noon and ride home in the afternoon or evening. This system was, until recently, still characteristic of European life.

The day's journey was also a limiting factor in visiting relatives and friends. Extending the journey beyond a day usually involved an expensive overnight stay in an inn. The importance of distance can be seen in its effect upon migration into London. The descendants of the people who lived within one day's journey to London were much more numerous than those from beyond this distance. This is clear from the carefully collected data of Professor Ekwall.[46] (Table 1).

---

[42] Fox [13].

[43] By including the cotter population the average size of family on a regular holding would increase from 3.5 to about 4.3 persons.

[44] Even a town of 200-400 apparently had to support part of its inhabitants by some sort of non-agricultural means.

[45] By law they were not allowed to be any closer together. Dickinson [8], pp. 41-42.

[46] Russell [24], a review of Ekwall [11].

TABLE 1

Distance from London	Number of migrants listed by Ekwall to each 1000 population in 1377.
Up to twenty miles	26.0
30 - 50 miles	4.8
60 - 100 miles	2.4
110 - 160 miles	1.4
180 miles and more	1.2

Of medieval migrants to Frankfurt-am-Main, a third came from within ten miles, nearly half between ten and fifty miles, and the rest from a greater distance.[47]

Most large cities needed access to both water and road transportation to obtain their supplies. But even with such access, the primitive transportation technology would tend to limit population. It is unlikely that an imperial Roman population of more than 300,000 could have been supplied by the modest network of roads leading into the city and the system of barges pulled against the Tiber current. Nor does it seem possible for a Cordoba of half a million[48] to have been supplied by local roads and the Guadalquivir River.

In view of the expense and difficulty of transport, medieval man had a surprisingly wide horizon. As Professor Stenton has written of English roads in the Middle Ages, "Every part of the country (was) within a fortnight's ride to London."[49] Even longer distance communication was simplified by concentrating essential activities within limited periods of time. The great fairs lasted only a few weeks. The Exchequer, the law courts and Parliament all held sessions in London at the same time in the fall.

In conclusion, the medieval region constituted a well organized web of settlements with important economic functions. Its pattern was determined by its rather low level of productivity and specialization and by the primitive transportation system. It often did constitute a political entity and less frequently a religious one. However, historians have tended to over-emphasize political and cultural aspects of medieval life, and in so doing have ignored other important aspects of regional structure.

---

[47] Bücker [5], p. 454. The Jews migrated further (p. 599).

[48] The size sometimes suggested. Sanches-Albornoz [26], p. 343.

[49] Stenton [28], esp. p. 20.

REFERENCES

[1] Bartlett, J.N., "The Expansion and Decline of York in the Later Middle Ages," *Economic History Review* 12 (1959) 17-33.

[2] Beresford, M.W., and J.K.S. St. Joseph, *Medieval England. An Aerial Survey* (Cambridge, 1958).

[3] Berry, B.J.L. and W.L. Garrison, "Alternate Explanations of Urban Rank-Size Relationships," *Annals of the Association of American Geographers* 48 (1958) pp. 83-91.

[4] Bracey, H.E., "A Rural Component of Centrality applied to Six Southern Counties in the United Kingdom," *Economic Geography* 32 (1956) 45.

[5] Bücker, K., *Die Bevölkerung von Frankfurt am Main im XIV und XV Jahrhundert* 47 (Tübingen, 1886).

[6] Carus-Wilson, E.M., "Evidences of Industrial Growth on some Fifteenth-Century Manors," *Economic History Review* 12 (1959) 190-205.

[7] Christaller, W., *Die Zentralen Orte in Suddeutschland* (Jena, 1935).

[8] Dickinson, R.E., *City Region and Regionalism* (London, 1947).

[9] Dickinson, R.E., "The Development and Distribution of the Medieval German Town," *Geography* 27 (1942) pp. 9-53.

[10] Dickinson, R.E., "The Morphology of the Medieval German Town," *Geographical Review* 35 (1945) 74-97.

[11] Ekwall, E., *Studies on the Population of Medieval London* (Stockholm: Kungl. Vitterhets Histories och Antikvitets Handlingar Fil.-Fil. Serien 2, 1956).

[12] Face, R.D., "The Vectuarii in the Overland Commerce between Champagne and Southern Europe," *Economic History Review* 12 (1959) 239-46.

[13] Fox, J.C., *The Royal Forests of England* (London, 1905).

[14] Holmes, U.T., Jr., and A.H. Schutz, *A History of the French Language* (New York, 1938).

[15] Jolliffe, J.E.A., *Pre-feudal England; the Jutes* (London, 1933).

[16] Lamond, E. (ed.) *Walter of Henley's Husbandry* (London, 1890).

[17] Lopez, R.S., "The Evolution of Land Transport in the Middle Ages," *Past and Present* no. 9 (1956) 17-29.

[18] Lösch, A., *The Economics of Location*, translated by W.H. Woglon with the assistance of W.F. Stolper (New Haven, 1954).

[19] Müller-Wille, *Die Ackerfluren in Landesteil Birkenfeld* (Bonn Diss., 1936).

[20] Philbrick, A.K., "Principles of Areal Functional Organization in Regional Human Geography," *Economic Geography* 33 (1957) 299-336.

[21] Rogers, J.E.T., *A History of Agriculture and Prices in England* (Oxford, 1866), I and IV.

[22] Russell, J.C., *British Medieval Population* (University of New Mexico Press, 1948).

[23] Russell, J.C., *Late Ancient and Medieval Population* (Philadelphia, 1958, *Transactions of the American Philosophical Society*).

[24] Russell, J.C., "Medieval Midland and Northern Migrants to London, 1100-1365," *Speculum* 34 (1959) 642.

[25] Russell, J.C., "The Persistence of Physical Type," *The Social Studies* 41 (1950) 210-13.

[26] Sanchez-Albornoz, C., La España Musulmana, (Buenos Aires, 1946) I.

[27] Shepherd, W., *Historical Atlas* (New York, 1923).

[28] Stenton, F.M., "The Road System of Medieval England," *Economic History Review* 7 (1936) 1-21.

[29] Stephenson, C., *Borough and Town* (Cambridge, Mass., 1933).

[30] Stewart, J.Q., "Demographic Gravitation, Evidence and Applications," *Sociometry* 11 (1948) 31-58.

[31] Stewart, J.Q. and W. Warntz, "Physics of Population Distribution," *Journal of Regional Science I* (1958) 99-123.

[32] Stubbs, W., *Selected Charters*, etc. (Oxford, 9th ed. rev. by H.W.C. Davis.)

[33] Tait, J., *The Medieval English Borough* (Manchester, 1936).

[34] Ullman, E., "A Theory of Location of Cities," *American Journal of Sociology* 46 (1941) 853-64.

[35] Victoria County History, *Leicestershire*, I.

[36] Villari, P., *The Barbarian Invasions of Italy* (London, 1902), II, 253-254.

[37] Wolff, P., *Les "Estimes" Toulousaines des XIVe et XVe siécles* (Toulouse, 1956).

[38] Zipf, G.K., *Human Behavior and the Principle of the Least Effort* (Cambridge, 1949).

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

SOME ASPECTS OF STATE TO STATE COMMODITY FLOWS IN THE UNITED STATES.†

by Stephen Spiegelglas*

It goes without saying that the economic activity in a country as large and as advanced as the United States will be characterized by a multiplicity of relationships between its geographically distributed parts. Historically, these parts have developed separately, and this circumstance accounts to a large extent for the differences in their present economic status and the role that they play in our economy today. Two other factors have also significantly contributed to state and regional economic differentiation. They are: (a) differences in the spatial distribution of resources, both material and human, leading to varying economies of scale, and (b) the location factor, leading to differences in transport costs.

Viewed in this fashion, the interregional division of labor in the United States does not differ conceptually from that division of labor which is generally considered to be the prime force behind the international exchange of goods and services. However, while international trade has by far more obstacles to overcome than interregional trade, it is international trade which has attracted more attention. In fact, it is really surprising how little we, in the United States, know about trade between the different parts of our country, as compared with what we know, or can easily find in official documents, about our trade relations with Canada, Japan, or Turkey.

---

*The author is Associate Professor of Business Administration, Boston University.

†The author wishes to acknowledge the valuable comments made by Professors Richard M. Clewett, Leon N. Moses, and James M. Patterson. Thanks are also due to Professor Harry R. Rymer and his staff at the Computing Center (Dearborn Observatory) of Northwestern University.

This paper tries to fill a small part of that gap. Its rather limited scope is to present certain relevant information on state-to-state[1] shipments in 1958.

By necessity, and not by choice, we utilize the only source of data on such shipments that is available, namely, the 1% sample of audited carload waybills of Class I railroads, published by the Interstate Commerce Commission. The utilization of these data becomes, with each passing year, more and more difficult to justify. Not only are the size of the sample and the nondisclosure rules a serious handicap, but in addition it is necessary to consider the fact that all railroads have accounted for only 46.4 per cent of the total of ton-miles shipped by all means of transport in this country in 1958.[2] All this requires us to be very cautious in generalizing from the data. Certain generalizations simply could not hold at all.

In studying the state-to-state flows, expressed either in terms of tons or ton-miles, it becomes immediately apparent that the I.C.C. sample has a strong inherent bias in favor of raw material shipments. Very many manufactured goods are not being shipped by rail at all, or only to a small extent. Hence, exports from a highly industrial state are likely to appear smaller than they actually are. Vice versa, exports of the raw material-producing states tend to be proportionately greater than their tabulated imports of manufactures. It follows that the totals of either exports or imports of the individual states, given in terms of tons or ton-miles, do not allow us a proper comparison of the "foreign trade" of the individual states.

Computing the balances of trade for the individual states on the basis of the ton or ton-mile flows, established from the I.C.C. reports, cannot be a meaningful undertaking. In fact, even if we had such complete data on all shipments by all means of transport, it is still debatable whether computing such balances would lead anywhere. One could object, for instance, to the use of the tonnage figures on the ground that since tons of various commodities represent different values, a deficit, established for any given state on the basis of such data, could actually turn into a substantial surplus, if dollar values of shipments were used. And it is definitely the latter result which is more meaningful. A similar objection could be voiced against computing balances in terms of ton-miles. Again, since we do not know what the values of these shipments were, we would be at a loss to interpret a positive or a negative balance of any particular state. An additional difficulty would no doubt arise in this case in connection with the interpretation of the role played by the distance component. All this points to an urgent need for inter-state flows expressed in terms of dollars.

---

[1] Forty-eight states and the District of Columbia will be discussed. Alaska and Hawaii are, on the other hand, not covered.

[2] It is important to note that in 1948 this percentage stood at 64.4. The sources are: I.C.C., *Volume of Intercity Freight Traffic, Public and Private, by Kinds of Transportation, 1939-49*, I.C.C. Statement No. 5046, Sept. 1950 (for 1948) and I.C.C., *73th Annual Report*, 1959 (for 1958).

Does all this discussion lead us therefore to conclude that the I.C.C.'s ton and ton-mile figures are of little or no use? Not at all; at least two interesting applications can be suggested.[3]

First, one could compute the sum-totals of exports *and* imports, on the basis of both types of data, for each state. This might tell us something about the magnitude of "foreign trade" of each state and thus reflect on its level of economic activity and its relative role in the national economy. It can actually be demonstrated that if all the states are ranked according to their sum-totals of their exports and imports, such a ranking order is highly correlated[4] with others in which the criteria used to judge the economic status of the individual states are: the size of the state population, the value added by manufacture, or personal income. In other words, the highly populous, highly developed, and rich states tend to be also important trading centers. We would, of course, expect that on an a priori grounds. The I.C.C. data demonstrate this to be a fact.

This outcome restores some of our lost confidence in the data. It may be also pointed out that most of our previous objections to the use of the ton and ton-mile data are definitely less valid in this case, since all the states are likely to be affected by the inadequacy of the data, some on the export side and others on the import side. It is possible that some industrial states will show relatively smaller sum-totals, if both their exports and imports, as is likely, consisted to a greater extent than those of the other states of manufactured goods, which were also shipped by trucks, by air, or otherwise. However, this should not be serious and the discrepancy could possibly be offset by the sheer magnitude of those states' sum-totals, thereby not affecting their position in the ranking order at all. The high correlation which is obtained demonstrates that the handicap arising from incomplete data is actually not serious in this case.

A second interesting application of the I.C.C. data consists in establishing for each state the average distances, in miles, of both its exports and imports. As shown in Figures 1 and 2, this information throws much light on the geographic distribution of markets for producers located in the individual states. Thus, we observe that the distances of exports exceeded those of imports for most of the western and southern states, while the opposite is generally true in the case of the states located in the eastern part of this country.

---

[3]Not discussed here, but definitely useful, is the detail on individual commodity flows contained in the I.C.C. statements.

[4]Spearman's coefficient of rank correlation is 0.83 if the states are ranked according to sum-totals, expressed in terms of tons, and 0.81 if ton-mile figures are utilized. The first result is obtained when we correlate columns (4) and (5) and the second when we correlate columns (4) and (6) given in Table 1.

TABLE 1: Selected Indicators of Economic Activity, by States, 1958

	Ranks according to:					
	Population (1)	Value Added (2)	Personal Income (3)	Average of columns (1) - (3) (4)#	Sum-Total of Exports and imports In tons (5)	In ton-miles (6)
1. Alabama	19	22	24	22	16	21
2. Arizona	35	40	34	37	41	37
3. Arkansas	32	35	36	34	30	36
4. California	2	2	2	2	10	1
5. Colorado	33	30	29	31	33	31
6. Connecticut	26	13	16	18.5	39	41
7. Delaware	46	39	42	41	44	45
8. District of Col.	40	44	35	39	47	48
9. Florida	12	24	12	15	12	14
10. Georgia	16	19	20	18.5	20	20
11. Idaho	43	41	44	43	35	32
12. Illinois	4	3	3	3	3	2
13. Indiana	10	8	10	10	11	15
14. Iowa	23	23	21	23.5	22	22
15. Kansas	29	29	26	27	24	23
16. Kentucky	21	21	25	23.5	9	10
17. Louisiana	20	25	23	25	17	19
18. Maine	36	33	38	36	38	39
19. Maryland	22	15	15	16	21	25
20. Massachusetts	9	9	9	9	32	27
21. Michigan	7	6	7	6	13	13
22. Minnesota	18	20	17	18.5	8	12
23. Mississippi	28	34	33	32	34	38
24. Missouri	13	12	11	11	15	16
25. Montana	42	45	41	43	37	35
26. Nebraska	34	36	32	33	27	30
27. Nevada	49	47	47	48	45	44
28. New Hampshire	45	37	46	43	48	47
29. New Jersey	8	7	8	8	18	17
30. New Mexico	39	43	39	40	40	33
31. New York	1	1	1	1	7	7
32. North Carolina	11	14	18	13	23	26
33. North Dakota	44	48.5	45	46	36	40
34. Ohio	6	5	5	5	2	4
35. Oklahoma	27	32	27	28.5	28	28
36. Oregon	31	27	28	28.5	26	9
37. Pennsylvania	3	4	4	4	1	3
38. Rhode Island	37	31	37	35	49	49
39. South Carolina	25	26	31	26	31	34
40. South Dakota	41	46	43	45	42	43
41. Tennessee	17	16	22	18.5	19	24
42. Texas	5	10	6	7	6	6
43. Utah	38	38	40	38	29	29
44. Vermont	47	42	49	47	46	46
45. Virginia	15	18	14	14	5	8
46. Washington	24	17	19	21	25	11
47. West Virginia	30	28	30	30	4	5
48. Wisconsin	14	11	13	12	14	18
49. Wyoming	48	48.5	48	49	43	42

It should be explicitly pointed out that the ranks in this column are not averages of columns (1) through (3), but rather reflect the order of the magnitudes of the 49 averages.

Source: Interstate Commerce Commission, Bureau of Transport Economics and Statistics, *State-to-State Distribution, All Commodities*, Carload Waybill Statistics for 1958, Statement SS-1, Washington, D.C., September 1959.

FIGURE 1: Miles of Haul of Exports (Adjusted) and Imports (Adjusted), by States, 1958. [Imports are shown in square brackets. The term "adjusted" means inclusive of intra-state hauls. Source: as in Table 1.]

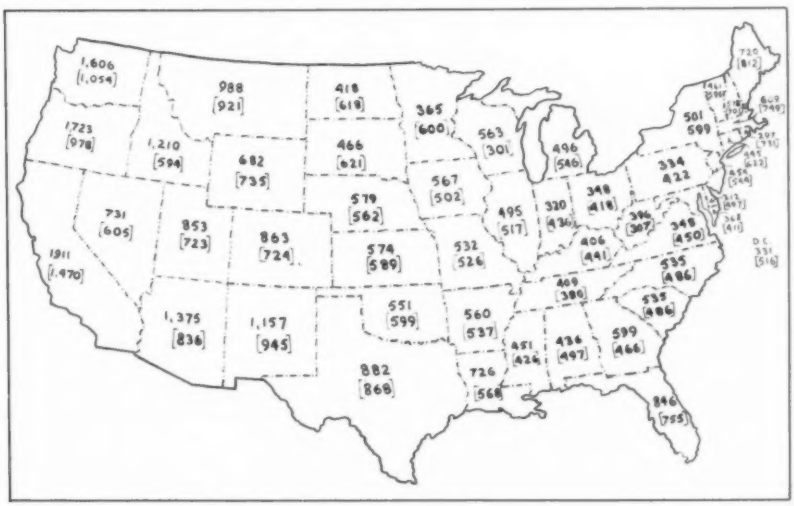


FIGURE 2: Miles of Haul of Exports and Imports, by States, 1958. [Imports are shown in square brackets. Source: as in Table 1.]

Off hand, one could argue that this result is quite inconsistent with what we are taught by the general location theory, namely, that the friction of distance should affect more the heavy and less expensive shipments of raw materials rather than the lighter and more expensive shipments of manufactures. Accordingly, we should expect, for instance, the exports from the raw material-producing states located in the western and southern part of this country to travel shorter distances than their imports, consisting primarily of manufactured commodities. However, these seemingly inconsistent results can be sufficiently well explained.

It is necessary to keep in mind, first of all, that the distance averages given for each state are definitely affected by the product-mix of both the exports and imports. More detailed analyses, such as Ullman's,[5] carried through for individual types of products[6] on the basis of the same data for certain previous years, have clearly demonstrated that differences in average haul for the various groups of products do actually exist.[7] Second, it can be shown statistically that a series in which states are ranked according to their average distance of haul is negatively correlated with another series in which the states are ranked according to their tonnage of shipment. This holds true for both exports and imports.[8] Third, our findings can be definitely reconciled with what we know about the patterns of population concentration, industrial location, and raw material sites in this country. Consideration of the population factor is particularly important, since it helps us to explain the generally smaller figures for the average haul for the states in the industrial belt and the longer average hauls for those on the periphery, as shown especially on Figure 2.

---

[5]Ullman, E.L., *American Commodity Flow*, (University of Washington Press, 1957), p. 29.

[6]The I.C.C. reports distinguish between five types of products, namely, agricultural, animal, mining, forest, and manufactured.

[7]An attempt to classify each state according to the highest percentage of its shipments in a particular type of product, has regrettably led to obviously confusing results. The tabulation, according to tonnage of exports, yielded the following breakdown: 21 mining states, 17 manufacturing, 7 agricultural, and 4 exporting predominantly forest products. However, such important industrial states as, for example, Illinois, Michigan, Ohio, and Pennsylvania appeared, on account of the previously noted incompleteness of the shipment data, among the typically mining states. In some cases, because of a 2 or 3-way tie, it was in addition very difficult, if not impossible, to determine where a given state should belong.

[8]Spearman's coefficients of rank correlation were as follows: for exports-adjusted -0.36; for exports -0.19; for imports adjusted -0.34; and for imports -0.50 (See Table 2).

TABLE 2: Ranking Order of States According to Miles and Tons, 1958

	Exports-adjusted* Miles	Rank	Exports Miles	Rank	Imports-adjusted* Miles	Rank	Imports Miles	Rank	RANKS according to tons: Exports-adjusted	Exports	Imports-adjusted	Imports
1. Alabama	285	39	436	35	275	41	497	36	13	10	20	23
2. Arizona	734	6	1,375	4	464	14	836	7	41	43	38	39
3. Arkansas	361	26	560	23	265	45	537	31	27	27	33	34
4. California	809	5	1,911	1	705	2	1,470	1	10	14	9	12
5. Colorado	431	13	863	9	393	23	724	13	30	37	32	31
6. Connecticut	304	35	445	34	566	7	622	16	44	46	35	27
7. Delaware	274	42	312	48	469	13	497	37	46	45	41	35
8. District of Col.	331	31	331	46	516	11	516	34	49	49	44	37
9. Florida	260	45	846	11	255	46	755	9	8	21	8	20
10. Georgia	330	32	599	17	316	31	466	39	18	23	15	18
11. Idaho	814	4	1,210	5	299	36	594	23	36	33	37	40
12. Illinois	274	41	495	29	306	33	517	33	3	5	3	2
13. Indiana	241	46	320	47	328	29	436	42	11	8	12	8
14. Iowa	376	23	567	21	331	28	502	35	20	18	23	22
15. Kansas	399	18	574	20	301	35	589	24	17	15	25	30
16. Kentucky	348	27	406	39	247	48	441	41	5	3	22	24
17. Louisiana	395	19	726	13	366	27	568	26	16	20	14	17
18. Maine	427	14	720	14	415	18	812	8	37	40	36	38
19. Maryland	323	33	368	41	382	24	411	46	28	22	21	11
20. Massachusetts	449	11	609	16	695	3	749	10	42	41	26	21
21. Michigan	295	36	496	28	371	26	546	29	12	12	11	9
22. Minnesota	217	49	365	42	268	43	600	20	6	6	7	16
23. Mississippi	287	38	451	33	255	47	426	43	32	32	34	33
24. Missouri	383	22	532	26	396	21	526	32	15	13	17	13
25. Montana	674	8	988	7	321	30	921	5	34	31	39	46
26. Nebraska	334	30	579	18	277	40	562	28	24	25	27	29
27. Nevada	718	7	731	12	568	6	605	19	43	42	49	48
28. New Hampshire	462	10	578	19	639	4	701	15	47	47	45	42
29. New Jersey	344	29	454	32	500	12	544	30	31	28	13	7
30. New Mexico	926	2	1,157	6	448	15	945	4	38	34	43	47
31. New York	385	21	501	27	525	10	599	22	14	9	6	5
32. North Carolina	273	44	535	25	380	25	486	38	29	36	19	14
33. North Dakota	365	24	418	36	401	20	618	18	33	29	40	41
34. Ohio	231	48	348	44	304	34	418	45	4	4	1	1
35. Oklahoma	346	28	551	24	288	38	599	21	23	24	28	32
36. Oregon	1,213	1	1,723	2	544	8	978	3	21	17	31	28
37. Pennsylvania	240	47	334	45	266	44	422	44	2	2	2	4
38. Rhode Island	289	37	297	49	722	1	731	12	48	48	48	44
39. South Carolina	273	43	410	37	417	17	564	27	35	35	29	26
40. South Dakota	390	20	466	30	395	22	621	17	40	39	42	45
41. Tennessee	278	40	409	38	283	39	380	47	19	16	18	15
42. Texas	406	17	882	8	447	16	868	6	7	11	5	10
43. Utah	435	12	853	10	273	42	723	14	26	30	30	36
44. Vermont	421	15	461	31	533	9	580	25	45	44	46	43
45. Virginia	322	34	348	43	403	19	450	40	9	7	4	3
46. Washington	875	3	1,606	3	621	5	1,054	2	25	26	24	25
47. West Virginia	363	25	396	40	208	49	307	49	1	1	16	19
48. Wisconsin	410	16	563	22	298	37	331	48	22	19	10	6
49. Wyoming	574	9	682	15	308	32	735	11	39	38	47	49

* Adjusted = inclusive of intra-state hauls.

Source: Interstate Commerce Commission, Bureau of Transport Economics and Statistics, *State-to-State Distribution, All Commodities*, Carload Waybill Statistics for 1958, Statement SS-1, Washington, D.C., September 1959.

As a corollary, we find that the average length of haul, computed for the country as a whole, was 363 miles in 1958. The corresponding figure for 1948 was 405 miles.[9] The 10 per cent decline[10] in a period of one decade suggests that the individual states, or probably more correctly stated - regions, are becoming more self-sufficient. This would support Ullman's contention that the United States "definitely is not one market for most producers."[11] Actually, this is what we should have expected all along. Indeed, little imagination is necessary to figure out that the spreading of the population all over this country, the rise in technology as well as its rapid dissemination, and the consequent increased market orientation of the industrial producers, would tend to shorten the average length of haul. The I.C.C. data seem to provide the much needed evidence on that point.

[9] Ullman, E.L., *op. cit.*, p. 29.

[10] Inasmuch as both mileage figures are subject to sampling errors, this 10 per cent decline must be interpreted with caution. It is the direction of change rather than its magnitude which is relevant.

[11] Ullman, E.L., *op. cit.*, p. 30. Professor Moses was kind enough to point out that supporting evidence for Ullman's and my arguments lies in the fact that we should have actually expected the railroads to have, on account of the increasing competition from the trucks, longer rather than shorter hauls. Since their average haul declined, some other cause must have been operating. The tendency toward regionalization of economic activity should be recognized as such a major cause.

APPENDIX

TABLE A: Tons and Ton-Miles of Haul, by States, 1958

	Tons		Ton - Miles	
	Exports-adjusted	Imports-adjusted	Exports-adjusted	Imports-adjusted
1. Alabama	292,425	235,026	83,321,400	64,563,200
2. Arizona	42,376	42,443	31,119,700	19,694,200
3. Arkansas	129,908	90,578	46,860,400	24,016,700
4. California	405,898	437,408	328,517,400	308,344,900
5. Colorado	98,899	106,142	42,584,300	41,750,900
6. Connecticut	17,330	62,839	5,272,900	35,574,800
7. Delaware	13,529	30,379	3,706,700	14,240,700
8. District of Col.	1,305	23,147	432,200	11,949,400
9. Florida	439,908	454,616	114,501,900	115,873,800
10. Georgia	207,307	269,093	68,450,100	84,925,400
11. Idaho	81,720	49,908	66,498,000	14,940,200
12. Illinois	865,316	955,677	237,287,100	292,484,700
13. Indiana	329,058	371,346	79,285,500	121,690,600
14. Iowa	194,764	187,217	73,315,300	61,866,600
15. Kansas	223,808	140,766	89,327,100	42,415,600
16. Kentucky	611,867	203,145	212,978,900	50,181,200
17. Louisiana	224,902	276,841	88,801,200	101,323,100
18. Maine	63,754	54,354	27,198,400	22,548,100
19. Maryland	116,036	208,804	37,418,400	79,727,000
20. Massachusetts	35,178	129,563	15,796,300	90,046,900
21. Michigan	294,373	377,345	86,831,500	139,857,600
22. Minnesota	561,374	462,609	121,959,900	123,983,500
23. Mississippi	98,056	87,921	28,090,900	22,399,200
24. Missouri	228,748	259,424	87,682,800	102,809,200
25. Montana	83,451	40,420	56,234,700	12,987,000
26. Nebraska	153,725	128,017	51,285,400	35,512,700
27. Nevada	25,182	7,068	18,090,600	4,014,100
28. New Hampshire	7,831	18,303	3,613,700	11,685,600
29. New Jersey	98,688	298,498	33,979,400	149,209,500
30. New Mexico	62,208	23,288	57,587,900	10,426,100
31. New York	272,428	542,688	104,988,100	284,882,500
32. North Carolina	108,442	239,251	29,605,800	91,015,700
33. North Dakota	83,455	36,437	30,469,000	14,614,000
34. Ohio	793,328	1,039,708	183,531,400	316,384,200
35. Oklahoma	158,668	115,294	54,931,100	33,211,100
36. Oregon	175,434	109,130	212,874,900	59,310,700
37. Pennsylvania	1,132,315	959,310	271,215,500	254,724,100
38. Rhode Island	5,656	13,864	1,633,600	10,012,500
39. South Carolina	81,854	114,598	22,359,600	47,750,600
40. South Dakota	46,899	24,236	18,283,800	9,569,200
41. Tennessee	202,848	248,460	56,472,700	70,233,700
42. Texas	516,656	572,570	209,716,800	256,042,200
43. Utah	147,294	111,722	64,040,800	30,500,700
44. Vermont	15,378	17,015	6,469,600	9,065,000
45. Virginia	407,911	664,883	131,201,000	268,011,200
46. Washington	152,810	161,687	133,660,300	100,342,000
47. West Virginia	1,140,257	263,645	413,915,100	54,871,300
48. Wisconsin	168,912	389,938	69,173,500	116,044,400
49. Wyoming	51,270	14,118	29,402,600	4,347,600

Adjusted = inclusive of intra-state hauls.

Source: Interstate Commerce Commission, Bureau of Transport Economics and Statistics, *State-to-State Distribution, All Commodities*, Carload Waybill Statistics for 1958, Statement SS-1, Washington, D.C., September 1959.

TABLE B: Tons and Ton-Miles of State-to-State Flows, 1958

	Tons		Ton - Miles	
	Exports	Imports	Exports	Imports
1. Alabama	160,075	102,676	69,806,600	51,048,400
2. Arizona	21,322	21,389	29,315,300	17,889,800
3. Arkansas	74,622	35,292	41,790,500	18,946,800
4. California	150,836	182,346	288,223,300	268,050,800
5. Colorado	43,812	51,055	37,797,600	36,964,200
6. Connecticut	11,237	56,746	5,005,300	35,307,200
7. Delaware	11,715	28,565	3,655,100	14,189,100
8. District of Col.	1,305	23,147	432,200	11,949,400
9. Florida	106,887	121,595	90,397,400	91,769,300
10. Georgia	91,851	153,637	55,039,300	71,514,600
11. Idaho	53,043	21,231	64,178,100	12,620,300
12. Illinois	382,655	473,016	189,547,400	244,745,000
13. Indiana	204,767	247,055	65,586,700	107,991,800
14. Iowa	117,667	110,120	66,677,800	55,229,100
15. Kansas	135,544	52,502	77,815,100	30,903,600
16. Kentucky	496,999	88,277	201,739,200	38,941,500
17. Louisiana	107,879	159,818	78,279,900	90,801 800
18. Maine	32,347	22,947	23,285,700	18,635,400
19. Maryland	98,563	191,331	36,309,400	78,618,000
20. Massachusetts	25,112	119,497	15,279,900	89,530,500
21. Michigan	155,612	238,584	77,251,800	130,277,900
22. Minnesota	260,814	162,049	95,250,500	97,274,100
23. Mississippi	53,590	43,455	24,192,100	18,500,400
24. Missouri	151,636	182,312	80,733,200	95,859,600
25. Montana	53,762	10,731	53,132,000	9,884,300
26. Nebraska	78,446	52,738	45,405,100	29,632,400
27. Nevada	24,712	6,598	18,065,100	3,988,600
28. New Hampshire	5,948	16,420	3,438,900	11,510,800
29. New Jersey	72,456	272,266	32,888,400	148,118,500
30. New Mexico	49,113	10,193	56,798,500	9,636,700
31. New York	184,929	455,189	92,586,600	272,481,000
32. North Carolina	44,486	175,295	23,795,600	85,205,500
33. North Dakota	65,830	18,812	27,487,300	11,632,300
34. Ohio	424,492	670,872	147,610,100	280,462,900
35. Oklahoma	89,351	45,977	49,253,400	27,533,400
36. Oregon	119,150	52,846	205,230,900	51,666,700
37. Pennsylvania	637,292	464,287	212,614,600	196,123,200
38. Rhode Island	5,490	13,698	1,629,800	10,008,700
39. South Carolina	45,106	77,850	18,501,300	43,892,300
40. South Dakota	34,699	12,036	16,182,600	7,468,000
41. Tennessee	121,199	166,811	49,589,700	63,350,700
42. Texas	157,490	213,404	138,909,100	185,234,500
43. Utah	60,207	24,635	51,342,800	17,802,700
44. Vermont	13,834	15,471	6,375,200	8,970,600
45. Virginia	207,931	464,903	72,432,200	209,242,400
46. Washington	77,324	86,201	124,206,500	90,888,200
47. West Virginia	1,015,251	138,639	401,639,800	42,596,000
48. Wisconsin	113,335	334,361	63,853,400	110,724,300
49. Wyoming	42,192	5,040	28,760,500	3,705,500

Source: As in Appendix Table A

JOURNAL OF REGIONAL SCIENCE, VOL. 2, NO. 2, 1960

A NOTE ON THE DERIVATION OF NEAREST NEIGHBOR DISTANCES.

by Michael F. Dacey*

1. INTRODUCTION

A derivation of the mean and variance of a class of nearest neighbor distances is given in this paper. A nearest neighbor measurement is the linear distance between a specified pair of nearest neighbors. The analysis of these distances contributes a high degree of objectivity to the description of physical, cultural, biological or abstract phenomena which have discrete spatial locations and may be mapped as points.

The central concept of nearest neighbor analysis is randomness. A punctiform distribution is random when each areal unit of the region containing the points has an equal probability of receiving a point. A point pattern that is not random is either more clustered than random or more uniform than random. Nearest neighbor analysis is a systematic method for reducing the many subtleties of point distributions to a succinct verbal description--a pattern is called random, more grouped than random, or more uniform than random.

Nearest neighbor relations are derived under the assumption that the points in an area are randomly distributed in accordance with the Poisson distribution. These relations, then, may be used to detect the presence of nonrandomness in a given pattern. The hypothesis of randomness may be tested in one of two ways. The distribution of the nearest neighbor distances may be regarded as approximately normal and, in this case, ordinary probability theory is applicable. Clark

*The author is Assistant Professor of Regional Science at the University of Pennsylvania.

and Evans[1] have suggested this approach. Alternatively, as Thompson[2] has demonstrated, the nearest neighbor distances are distributed as chi-square so that the presence of nonrandomness in a distribution may be detected by a chi-square test.

Both tests are relatively effective for determining the presence of nonrandomness in a distribution. Because a test is able to detect nonrandomness, it does not follow, however, that the nature of the nonrandomness may be determined by these tests. A nearest neighbor method is not effective, and should not be used, to test the hypothesis that a distribution has a systematic, nonrandom pattern, such as grouped or uniform. As a consequence, while a distribution may be called random, or said to deviate positively or negatively from random, it can not be said on the basis of this that a distribution is uniform or grouped. As such, nearest neighbor methods are appropriate when there is reason to believe a point distribution has been generated by random variables. If there is reason to believe that the factors generating the point distribution are not random variables, as is the case with the hexagonal distribution of cities (which may be represented by points) derived from central place theory, nearest neighbor methods are not applicable.

Nearest neighbor methods were conceived and largely developed by statistical ecologists for the description of phytosociological distributions.[3] The present study is based upon this work, and the following derivation of nearest neighbor distances resembles those by Morisita[4] and Clark and Evans, but it is more general.

2. DEFINITIONS

Assume an infinite number of points are located in an unbounded area such that the density of points per unit area is d. This distribution is random if the points are located such that each areal unit of the region has an equal probability of receiving a point. When the area π contains a population of randomly distributed individuals with density $m = d\pi$, it is generally accepted that the probability that a randomly chosen area within the specified region contains exactly x points is given by the Poisson distribution and is $\frac{m^x e^{-m}}{x!}$. Let the specified area be a sector of a circle of radius r formed by dividing the circle into K equal size sectors.

---

[1]Clark, P. J. and F. C. Evans, "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations." *Ecology*, 35 (1954), 445-453.

[2]Thompson, H. R. "Distribution of Distance to *n*th Neighbor in a Population of Randomly Distributed Individuals." *Ecology*, 37 (1956), 391-394.

[3]Statistical methods used by ecologists are reviewed in P. Greig-Smith, *Quantitative Plant Ecology*, New York (1957).

[4]Morisita, M. "Estimation of Population Density by Spacing Method," *Memoirs of the Faculty of Science*, Kyushu University, Series E-1 (1954), 187-197.

Given any center point i, its nearest neighbor within the kth of K sectors, where $k = 1, 2, \ldots, K$, is called $j(k/K)$. The identification of both the neighbors and the sectors is determined by the rank order of the linear distances from i to these $j(k/K)$ neighbors. The sector $k = 1$ contains the nearest neighbor of i, and this neighbor is called $j(1/K)$. The sector $k = 2$ contains the second nearest $j(k/K)$ neighbor, and it is called $j(2/K)$. Similarly, the $k = K$ sector contains the most distant $j(k/K)$ neighbor, and it is called $j(K/K)$. These relations are illustrated in Figure 1 for $K = 6$.

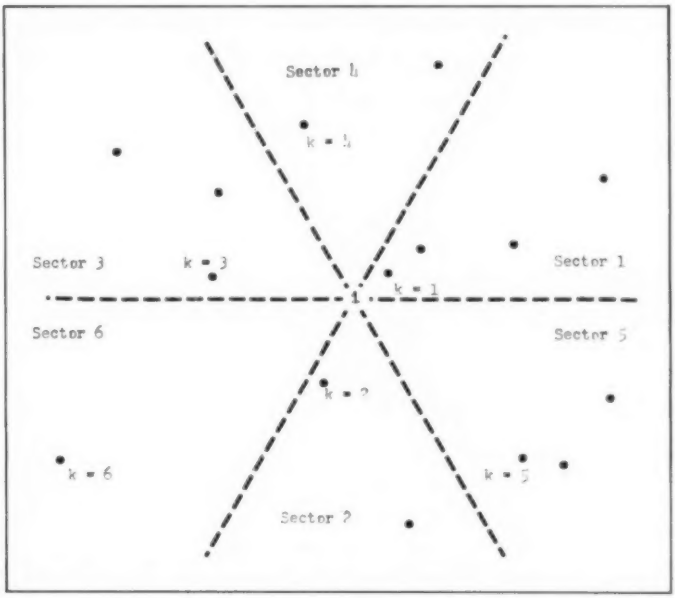


FIGURE 1: Illustration of the nearest neighbor relations used to identify the k sectors and the j_k neighbors when $K = 6$. The sextant is centered on the i point, the nearest point in each sector is determined, and the distance is measured from the i point to these points. The sector k_1 contains the nearest specified point, called j_1; the sector k_2 contains the second nearest specified point, called j_2; and the identification continues in a similar manner so that the sector k_6 contains the most distant of the six specified points, called j_6.

3. DERIVATION OF THE PROBABILITY DISTRIBUTIONS

Since m is the mean density of points in the area π, $\frac{m}{K}$ is the mean number of points per sector. It follows that the probability of finding x points in a sector of $\frac{\pi r^2}{K}$ units is $\frac{[(m/K)r^2]^x e^{-(m/K)r^2}}{x!}$. As the probability of not finding at least one point within the distance r of the point i is $q = e^{-(m/K)r^2}$, and since $p = 1 - q$, the probability of finding at least one point within the distance r is $p = 1 - e^{-(m/K)r^2}$.

The probability of finding one point in each of the sectors is then seen to be $P_{j(k/K)} = C_{kK} p^k q^{(K-k)}$, where the C_{kK} denotes the combinations of the K sectors taken k at a time. Let $z = -(m/K)r^2$. Then, since $p^k = 1 - e^{zk}$, and $q^{(K-k)} = e^{z(K-k)}$, we obtain:

$$P_{j(0/K)} = e^{zK}$$

$$P_{j(1/K)} = K[e^{z(K-1)} - e^{zK}]$$

$$P_{j(2/K)} = C_{2K}[e^{z(K-2)} - 2e^{z(K-1)} + e^{zK}]$$

$$P_{j(3/K)} = C_{4K}[e^{z(K-3)} - 3e^{z(K-2)} + 3e^{z(K-1)} - e^{zK}]$$

. . .
. . .
. . .

The probability that the $j(k/K)$ point is within the distance r of the center point i is:

$$P_{j(k/K)}(0 \sim r) = 1 - [P_{j(0/K)} + P_{j(1/K)} + \cdots + P_{j(k-1/K)}] .$$

The probability that the $j(k/K)$ point is between the distances r' and r from the point i is:

$$P_{j(k/K)}(r' \sim r) = P_{j(k/K)}(0 \sim r) - P_{j(k/K)}(0 \sim r') .$$

Let $z' = -(m/K)r'^2$, and these prove to be:

$$P_{j(1/K)}(r' \sim r) = e^{z'K} - e^{zK}$$

$$P_{j(2/K)}(r' \sim r) = K[e^{z'(K-1)} - e^{z(K-1)}] - [K-1][e^{z'K} - e^{zK}]$$

$$P_{j(3/K)}(r' \sim r) = C_{3K}[e^{z'(K-2)} - e^{z(K-2)}] - [2C_{3K} - K]$$

$$[e^{z'(K-1)} - e^{z(K-1)}] + [C_{3K} - K + 1]$$

$$[e^{z'K} - e^{zK}]$$

$$\begin{array}{ccc} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}$$

$$P_{j(k/K)}(r' \sim r) = C_{kK}[e^{z'(K-k+1)} - e^{z(K-k+1)}] - [(k-1)C_{kK} - C_{(k-1)K}]$$

$$[e^{z'(K-k+2)} - e^{z(K-k+2)}] + \ldots [C_{kK} - C_{(k-1)K}$$

$$+ \ldots 1] \; [e^{z'K} - e^{zK}]$$

The probability distribution of r is obtained by differentiating these last expressions with respect to r. These are:

$$F\,r_{j(1/K)} = 2mre^{zK}\Delta r$$

$$F\,r_{j(2/K)} = 2(K-1)mr\;[e^{z(K-1)} - e^{zK}]\;\Delta r$$

$$F\,r_{j(3/K)} = 2mr[C_{3K}(K-2)K^{-1}e^{z(K-2)} - (2C_{3K} - K)$$

$$(K-1)K^{-1}e^{z(K-1)} + (C_{3K} - K + 1)e^{zK}]\,\Delta r$$

$$\begin{array}{ccc} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}$$

$$F\,r_{j(k/K)} = 2mr[C_{kK}(K-k+1)K^{-1}e^{z(K-k+1)} - [(k-1)C_{kK} - C_{(k-1)K}]$$

$$(K-k+2)K^{-1}e^{z(k-k+2)} + \ldots (C_{kK} - C_{(k-1)K} + \ldots 1)$$

$$(e^{zK})]\,\Delta r$$

4. FIRST MOMENT

The first moment of r is obtained by multiplying each probability distribution by r and then integrating with respect to r over the interval from 0 to $+\infty$. These can be shown to be:

$$\mu_1' j(1/K) = (1/2)\sqrt{\pi/m}$$

$$\mu_1' j(2/K) = [K\sqrt{K/(K-1)} - (K-1)]\sqrt{\pi/m\,(2^{-1})}$$

$$\mu_1' j(3/K) = [C_{3K}\sqrt{K/(K-2)} - (2C_{3K} - K)\sqrt{K/(K-1)} + (C_{3K} - K + 1)]\sqrt{\pi/m\,(2^{-1})}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\mu_1' j(k/K) = [C_{kK}\sqrt{K/(K-k+1)} - [(k-1)C_{kK} - C_{(k-1)K}]\sqrt{K/(K-k+2)} + \ldots (C_{kK} - C_{(K-1)K} + \ldots 1]\sqrt{\pi/m\,(2^{-1})}$$

The expected mean distances between randomly distributed nearest neighbors are listed in Table 1.

TABLE 1: The expected mean distance, $E_{rk/K}$, to the j_k neighbor in the kth of K sectors in a random point distribution. The d is the density of points per unit area.

k	$E_{rk/1}$	$E_{rk/2}$	$E_{rk/3}$	$E_{rk/4}$	$E_{rk/5}$	$E_{rk/6}$
1	$\frac{.5000}{\sqrt{d}}$	$\frac{.5000}{\sqrt{d}}$	$\frac{.5000}{\sqrt{d}}$	$\frac{.5000}{\sqrt{d}}$	$\frac{.5000}{\sqrt{d}}$	$\frac{.5000}{\sqrt{d}}$
2		$\frac{.9142}{\sqrt{d}}$	$\frac{.8371}{\sqrt{d}}$	$\frac{.8095}{\sqrt{d}}$	$\frac{.7951}{\sqrt{d}}$	$\frac{.7863}{\sqrt{d}}$
3			$\frac{1.2610}{\sqrt{d}}$	$\frac{1.1237}{\sqrt{d}}$	$\frac{1.0698}{\sqrt{d}}$	$\frac{1.0403}{\sqrt{d}}$
4				$\frac{1.5669}{\sqrt{d}}$	$\frac{1.3809}{\sqrt{d}}$	$\frac{1.3034}{\sqrt{d}}$
5					$\frac{1.8444}{\sqrt{d}}$	$\frac{1.6175}{\sqrt{d}}$
6						$\frac{2.1010}{\sqrt{d}}$

5. SECOND MOMENT

The second moment of r is obtained by multiplying each of the probability distributions by r^2 and then integrating with respect to r over the interval from 0 to $+\infty$. These can be shown to be:

$$\mu_2' j(1/K) = 1/m$$

$$\mu_2' j(2/K) = [K^2(K-1)^{-1} - (K-1)]\ /m$$

$$\mu_2' j(3/K) = [KC_{3K}(K-2)^{-1} - K(2C_{3K} - K)(K-1)^{-1} + (C_{3K} - K + 1)]\ /m$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\mu_2' j(k/K) = [KC_{kK}(K-k+1)^{-1} - K[(K-k-1)C_{kK} - C_{(k-1)K}](K-k+2)^{-1}$$

$$+ \ldots (C_{kK} - C_{(k-1)K} + \ldots 1)]\ /m$$

6. VARIANCE

The variance of a nearest neighbor distance is obtained from $\sigma^2_{r(k/K)} = \mu_2' j(k/K) - [\mu_1' j(k/K)]^2$. The variance is given in Table 2 for all $K \leq 6$.

TABLE 2: The variance, $\sigma^2_{rk/K}$, of the $E_{rk/K}$ in a random point distribution. The d is the density of points per unit area.

k	$\sigma^2_{rk/1}$	$\sigma^2_{rk/2}$	$\sigma^2_{rk/3}$	$\sigma^2_{rk/4}$	$\sigma^2_{rk/5}$	$\sigma^2_{rk/6}$
1	$\frac{.2146}{d}$	$\frac{.2146}{d}$	$\frac{.2146}{d}$	$\frac{.2146}{d}$	$\frac{.2146}{d}$	$\frac{.2146}{d}$
2		$\frac{.3744}{d}$	$\frac{.2986}{d}$	$\frac{.2747}{d}$	$\frac{.2640}{d}$	$\frac{.2577}{d}$
3			$\frac{.5045}{d}$	$\frac{.3665}{d}$	$\frac{.3212}{d}$	$\frac{.3007}{d}$
4				$\frac{.6202}{d}$	$\frac{.4260}{d}$	$\frac{.3629}{d}$
5					$\frac{.7296}{d}$	$\frac{.4806}{d}$
6						$\frac{.8324}{d}$

JOURNAL OF REGIONAL SCIENCE

VOLUMES I AND II

Edited by W. Isard and M. B. Teitz

Managing Editor, B. H. Stevens

CONTENTS

ARTICLES:

Community Income Multipliers: A Population Growth Model	C. M. Tiebout	II, 1, 75-84
Commuting and the Labor Market Area	R. Gerard	I, 1, 124-130
Economic Structure of Regions Based on the Concept of Distance, A Comparison of the	R. Stone	II, 2, 1-20
General Interregional Equilibrium	W. Isard and D. Ostroff	II, 1, 67-74
Industrial Complex Analysis, Agglomeration Economies and Regional Development	W. Isard and E.W.Schooler	I, 2, 19-33
Inter-Areal Linear Programming for a Metropolitan Region, A Note on	B. H. Stevens and R.E. Coughlin	I, 2, 75-83
Interregional Linear Programming: An Elementary Presentation and a General Model	W. Isard	I, 1, 1-59
Interregional Linear Programming Model, An	B. H. Stevens	I, 1, 60-98
Intervening Opportunities and Competing Migrants	S. A. Stouffer	II, 1, 1-26
Intra-City Locational and Residential Patterns, The Measurement of	B. Duncan and O.D. Duncan	II, 2, 37-54
Location of U.S. Manufacturing, Changes in since 1929	V. R. Fuchs	I, 2, 1-17
Metropolitan City Region of the Middle Ages, The	J. Russell	II, 2, 55-70

CONTENTS (Continued)

ARTICLES:

Migration, Real Income and Information	P. Nelson	I, 2, 43-73
Nearest Neighbor Distances, A Note on the Derivation of	M. F. Dacey	II, 2, 81-87
Physics of Population Distribution	J.Q. Stewart and W. Warntz	I, 1, 99-123
Residential Activity in Urban Areas, A Model for the Distribution of	J. D. Herbert and B.H. Stevens	II, 2, 21-36
State-to-State Commodity Flows in the U.S., Some Aspects of	S. Spiegelglas	II, 2, 71-80
Statistical Methodology for Areal Distributions, Contributions to	W. Warntz and D. Neft	II, 1, 47-66
Transportation and Optimal Network Construction, Models of	R. E. Quandt	II, 1, 27-45

COMMENTS:

On 'Changes in the Location of U.S. Industry,' and 'Industrial Complex Analysis.'	P.G. Craig, D.Netzer and W. R. Thompson	I, 2,35-42
On 'Migration, Real Income and Information.'	W. Isard and M. B. Teitz	I, 2, 74

CONTRIBUTORS:

Coughlin, R. E.	I, 2, 75
Craig, P. G.	I, 2, 35
Dacey, M. F.	II, 2, 81
Duncan, B.	II, 2, 37
Duncan, O. D.	II, 2, 37
Fuchs, V. R.	I, 2, 1
Gerard, R.	I, 1, 124
Herbert, J. D.	II, 2, 21
Isard, W.	I, 1, 1
	I, 2, 19
	I, 2, 74
	II, 1,67
Neft, D.	II, 1, 47
Nelson, P.	I, 2, 43
Netzer, D.	I, 2, 37
Ostroff, D. J.	II, 1, 67
Quandt, R. E	II, 1, 27
Russell, J.	II, 2, 55
Schooler, E. W.	I, 2, 19
Spiegelglas, S.	II, 2, 71
Stone, R.	II, 2, 1
Stouffer, S. A.	II, 1, 1
Stewart, J. Q.	I, 1, 99
Stevens, B. H.	I, 1, 60
	I, 2, 75
	II, 2, 21
Teitz, M. B.	I, 2, 74
Thompson, W. R.	I, 2, 40
Tiebout, C. M.	II, 1, 75
Warntz, W.	I, 1, 99
	II, 1, 47

GRADUATE FELLOWSHIPS IN REGIONAL SCIENCE

Students who are interested in the analysis of economic, social, and political phenomena in a spatial and regional context are invited to apply for graduate fellowships in the Department of Regional Science at the University of Pennsylvania for the 1961-62 academic year. Awards up to $3500 per year plus tuition are available to those wishing to undertake a program of study leading to an M.A. or Ph.D. in Regional Science, and who are qualified applicants for admission to the Graduate School of Arts and Sciences. Candidates should have undergraduate preparation in one of the social sciences or in mathematics and statistics. Desirable backgrounds include city planning, geography, economics, statistics, mathematics, engineering, sociology, and political science.

Closing date for applications is February 17, 1961. Address inquiries to Professor Walter Isard, Chairman, Department of Regional Science, Wharton School, University of Pennsylvania, Philadelphia 4, Pa.

RESOLUT

100

INSTRUCTIONS Resolution is expr
lar film under specified conditions. Numerals
adjacent "L-shaped" groupings.

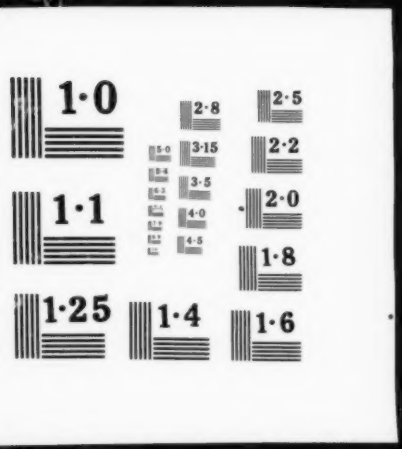
In microfilming it is necessary to determine t
group in which it can be decerned that the l
the reduction ratio the line below is 100 milli
dividing the length into 100 gives the reduction
and 100/10 = 10.

Examine the "L-shaped" line groupings in
adjacent to the group in which it is possible
only necessary to determine that there are line
these conditions it is considered that the group
discernible horizontally and vertically while th
is 10 and 7.9 × 10 = 79 lines per millimeter
are not recorded satisfactorily. Under the partic
per millimeter.

In order to check the camera for vibrati
of the field in one orientation and then turn
If the lines in both directions at both orientat

Resolution as measured on the film, is a te
processing, and other factors. It is recomme
film, at the beginning of the roll or preferabl
venient means of continually checking both t
referred to at any time.

TION CHART



100 MILLIMETERS

n is expressed in terms of the lines per millimeter recorded by a particu-
Numerals in this chart indicate the number of lines per millimeter in

termine the reduction ratio and multiply by the number adjacent to the
nat the lines are either horizontal or vertical. As an aid in determining
100 millimeters in length. Measuring this line in the film image and
eduction ratio. Example: the line is 10 millimeters long in the film image

pings in the film with a forty power microscope and note the number
possible to decern that the lines are either horizontal or vertical. It is
e are lines in both directions regardless of the number of lines. Under
he group in question is resolved. Example: the lines in 7.9 group are
while the lines in the 8.9 group are not discernible. Reduction ratio
llimeter recorded satisfactorily. 8.9 × 10 = 89 lines per millimeter which
he particular conditions maximum resolution is between 79 and 89 lines

r vibration several exposures should be made with the chart in the center
en turn it through an angle of 45° and make several more exposures.
orientations are equally resolved the camera is free from vibrations.

m, is a test of the entire photographic system, including lens, exposure,
recommended the resolution charts be placed on every roll of micro-
preferably at both ends of the roll. This practice will provide a con-
g both the operation and performance of the system which can be